PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12Q 1/68, C12N 9/14, 9/16

A1

(11) International Publication Number: WO 97/46701

(43) International Publication Date: 11 December 1997 (11.12.97)

(21) International Application Number: PCT/US97/08705

(22) International Filing Date: 20 May 1997 (20.05.97)

(30) Priority Data:
08/658,322 5 June 1996 (05.06.96) US
08/803,621 21 February 1997 (21.02.97) US

(71) Applicant: FOX CHASE CANCER CENTER [US/US]; 7701 Burholme Avenue, Philadelphia, PA 19111 (US).

(72) Inventor: YEUNG, Anthony, T.; 217 Walnut Place, Philadelphia, PA 19083 (US).

(74) Agents: HAGAN, Patrick, J. et al.; Dann, Dorfman, Herrell and Skillman, Suite 720, 1601 Market Street, Philadelphia, PA 19103-2307 (US).

(81) Designated States: AU, CA, JP, MX, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: MISMATCH ENDONUCLEASES AND USES THEREOF IN IDENTIFYING MUTATIONS IN TARGETED POLYNUCLEOTIDE STRANDS

(57) Abstract

An endonuclease isolated from celery, CEL I, is disclosed as well as methods for use in detection of mutations in targeted polynucleotides. The methods facilitate localization and identification of mutations, mismatches and polymorphisms. The enzyme recognizes every type of mismatch regardless of the sequence context in which the mismatch resides and the enzyme is active in pH ranges from acidic to basic.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

MISMATCH ENDONUCLEASES AND USES THEREOF IN IDENTIFYING MUTATIONS IN TARGETED POLYNUCLEOTIDE STRANDS

Pursuant to 35 U.S.C. §202(c), it is hereby acknowledged that the U.S. Government has certain rights in the invention described herein, which was made in part with funds from the National Institutes of Health, National Cancer Institute.

This application is a continuation-in-part of U.S. Application, Serial No. 08/658,322, filed June 5, 1996, the entire disclosure of which is incorporated by reference herein.

FIELD OF THE INVENTION

This invention relates to materials and methods for the detection of mutations in targeted nucleic acids. More specifically, the invention provides novel mismatch specific nucleases and methods of use of the enzyme that facilitate the genetic screening of hereditary diseases and cancer. The method is also useful for the detection of genetic polymorphisms.

BACKGROUND OF THE INVENTION

Several publications are referenced in this application by numerals in parenthesis in order to more fully describe the state of the art to which this invention pertains. Full citations for these references are found at the end of the specification. The disclosure of each of these publications is incorporated reference in the present specification.

The sequence of nucleotides within a gene can be mutationally altered or "mismatched" in any of several ways, the most frequent of which being base-pair substitutions, frame-shift mutations and deletions or insertions. These mutations can be induced by environmental factors, such as radiation

and mutagenic chemicals; errors are also occasionally committed by DNA polymerases during replication. Many human disease states arise because fidelity of DNA replication is not maintained. Cystic fibrosis, sickle cell anemia and some cancers are caused by single base changes in the DNA resulting in the synthesis of aberrant or non-functional proteins.

The high growth rate of plants and the abundance of DNA intercalators in plants suggests an enhanced propensity for mismatch and frameshift lesions. Plants and fungi are known to possess an abundance of single-stranded specific nucleases that attack both DNA and RNA (9-14). Some of these, like the Nuclease α of *Ustilago maydis*, are suggested to take part in gene conversion during DNA recombination (15,16). Of these nucleases, S1 nuclease from *Aspergillus oryzue* (17), and P1 nuclease from *Penicillium citrinum* (18), and Mung Bean Nuclease from the sprouts of *Vigna radiata* (19-22) are the best characterized. S1, P1 and the Mung Bean Nuclease are Zn proteins active mainly near pH 5.0 while Nuclease α is active at pH 8.0. The single strandedness property of DNA lesions appears to have been used by a plant enzyme, SP nuclease, for bulky adduct repair. The nuclease SP, purified from spinach, is a single-stranded DNase, an RNase, and able to incise DNA at TC_{6-4} dimers and cisplatin lesions, all at neutral pH (23,24). It is not yet known whether SP can incise DNA at mismatches.

In *Escherichia coli,* lesions of base-substitution and unpaired DNA loops are repaired by a methylation-directed long patch repair system. The proteins in this multienzyme system include MutH, MutL and MutS (1, 2). This system is efficient, but the C/C lesion and DNA loops larger than 4 nucleotides are not repaired. The MutS and MutL proteins are

conserved from bacteria to humans, and appear to be able to perform similar repair roles in higher organisms. For some of the lesions not well repaired by the MutS/MutL system, and for gene conversion where short-patch repair systems may be more desirable, other mismatch repair systems with novel capabilities are needed.

Currently, the most direct method for mutational analysis is DNA sequencing, however it is also the most labor intensive and expensive. It is usually not practical to sequence all potentially relevant regions of every experimental sample. Instead some type of preliminary screening method is commonly used to identify and target for sequencing only those samples that contain mutations. Single stranded conformational polymorphism (SSCP) is a widely used screening method based on mobility differences between single-stranded wild type and mutant sequences on native polyacrylamide gels. Other methods are based on mobility differences in wild type/mutant heteroduplexes (compared to control homoduplexes) on native gels (heteroduplex analysis) or denaturing gels (denaturing gradient gel electrophoresis). While sample preparation is relatively easy in these assays, very exacting conditions for electrophoresis are required to generate the often subtle mobility differences that form the basis for identifying the targets that contain mutations. Another critical parameter is the size of the target region being screened. In general, SSCP is used to screen target regions no longer than about 200-300 bases. The reliability of SSCP for detecting single-base mutations is somewhat uncertain but is probably in the 70-90% range for targets less than 200 bases. As the size of the target region increases, the detection rate declines, for example in

one study from 87% for 183 bp targets to 57% for targets 307 bp in length (35). The ability to screen longer regions in a single step would enhance the utility of any mutation screening method.

Another type of screening technique currently in use is based on cleavage of unpaired bases in heteroduplexes formed between wild type probes hybridized to experimental targets containing point mutations. The cleavage products are also analyzed by gel electrophoresis, as subfragments generated by cleavage of the probe at a mismatch generally differ significantly in size from full length, uncleaved probe and are easily detected with a standard gel system. Mismatch cleavage has been effected either chemically (osmium tetroxide, hydroxylamine) or with a less toxic, enzymatic alternative, using RNase A. The RNase A cleavage assay has also been used, although much less frequently, to screen for mutations in endogenous mRNA targets for detecting mutations in DNA targets amplified by PCR. A mutation detection rate of over 50% was reported for the original RNase screening method (36).

A newer method to detect mutations in DNA relies on DNA ligase which covalently joins two adjacent oligonucleotides which are hybridized on a complementary target nucleic acid. The mismatch must occur at the site of ligation. As with other methods that rely on oligonucleotides, salt concentration and temperature at hybridization are crucial. Another consideration is the amount of enzyme added relative to the DNA concentration.

The methods mentioned above cannot reliably detect a base change in a nucleic acid which is contaminated with more than 80% of a background nucleic acid, such as normal or wild type sequences.

Contamination problems are significant in cancer detection wherein a malignant cell, in circulation for example, is present in extremely low amounts. The methods now in use lack adequate sensitivity to be practically applied in the clinical setting.

A method for the detection of gene mutations with mismatch repair enzymes has been described by Lu-Chang and Hsu. See WO 93/20233. The product of the MutY gene which recognizes mispaired A/G residues is employed in conjunction with another enzyme described in the reference as an "all type enzyme" which can nick at all base pair mismatches. The enzyme does not detect insertions and deletions. Also, the all type enzyme recognizes different mismatches with differing efficiencies and its activity can be adversely affected by flanking DNA sequences. This method therefore relies on a cocktail of mismatch repair enzymes and DNA glycosylases to detect the variety of mutations that can occur in a given DNA molecule.

Often, in the clinical setting, the nature of the mutation or mismatch is unknown so that the use of specific DNA glycosylases is precluded. Thus, there is a need for a single enzyme system that is capable of recognizing all mismatches with equal efficiency and also detecting insertions and deletions, regardless of the flanking DNA sequences. It would be beneficial to have a sensitive and accurate assay for detecting single base pair mismatches which does not require a large amount of sample, does not require the use of toxic chemicals, is neither labor intensive nor expensive and is capable of detecting not only mismatches but deletions and insertions of DNA as well.

Such a system, coupled with a method that would facilitate the identification of the location of the mutation in a given DNA molecule would be clearly

advantageous for genetic screening applications. It is the purpose of the present invention to provide this novel mutation detection system.

SUMMARY OF THE INVENTION

The present invention provides materials and methods for the detection of mutations or mismatches in a targeted polynucleotide strand. Detection is achieved using novel endonucleases in combination with a gel assay system that facilitates the screening and identification of altered base pairing in targeted nucleic acid strands.

According to one aspect of the invention, there is provided a novel plant-based nuclease which is useful in the detection of mutations or mismatches in target DNA or RNA. Celery, for example, (*Apium graveolens var. dulce*) contains abundant amounts of the nuclease of the invention which is highly specific for insertional/deletional DNA loop lesions and mismatches. This enzyme, designated herein as CEL I, incises at the phosphodiester bond at the 3' side of the mismatched nucleotide. CEL I has been purified about 10,000 fold, so as to be substantially homogeneous.

In a preferred embodiment of th invention, a method is provided for determining a mutation in a target sequence of single stranded mammalian polynucleotide with reference to a non-mutated sequence of a polynucleotide that is hybridizable with the polynucleotide including the target sequence. The sequences are amplified by polymerase chain reaction (PCR), labeled with a detectable marker, hybridized to one another, exposed to CEL I of the present invention, and analyzed on gels for the presence of the mutation.

The plant-based endonuclease of the

invention has a unique combination of properties. These include the ability to detect all possible mismatches between the hybridized sequences formed in performing the method of the invention; recognize polynucleotide loops and insertions between such hybridized sequences; detect polymorphisms between such hybridized strands; recognize sequence differences in polynucleotide strands between about 100 bp and 3 kb in length and recognize such mutations in a target polynucleotide sequence without substantial adverse effects of flanking DNA sequences.

The plant-based endonuclease, CEL I of the invention is not unique to celery. Functionally similar enzymatic activities have been demonstrated in fourteen different plant species. Therefore, the enzyme is likely to be conserved in the plant kingdom and may be purified from plants other than celery. The procedure to purify this endonuclease activity from a plant other than celery is well known to those skilled in the art and is contemplated to be within the scope of the present invention. Such enzymes have been purified to substantial homogeneity from the plant species *Arabidopsis thaliana*, for example, in accordance with the present invention. This novel enzyme, designated ARA I, is like CEL I in its enzymatic activities and thus may be used to advantage in the genetic mutation screening assays of the invention.

The plant-based endonuclease may not be limited to the plant kingdom but may be found in other life forms as well. Such enzymes may serve functions similar to that of CEL I in celery or be adapted for other special steps of DNA metabolism. Such enzymes or the genes encoding them may be used or modified to produce enzymatic activities that can function the same or similar to CEL I. The isolation of such genes

and their modification is also within the scope of the present invention.

In another embodiment of the invention, the above-described method is employed in conjunction with the addition of DNA ligase, DNA polymerase or a combination thereof thereby reducing non-specific DNA cleavage.

In yet another embodiment of the invention, the simultaneous analysis of multiple samples is performed using the above-described enzyme and method of the invention by a technique referred to herein as multiplex analysis.

In order to more clearly set forth the parameters of the present invention, the following definitions are provided:

The term "endonuclease" refers to an enzyme that can cleave DNA internally.

The term "isolated nucleic acid" refers to a DNA or RNA molecule that is separated from sequences with which it is normally immediately contiguous (in the 5' and 3' directions) in the naturally occurring genome of the organism in which it originates.

The term "base pair mismatch" indicates a base pair combination that generally does not form in nucleic acids according to Watson and Crick base pairing rules. For example, when dealing with the bases commonly found in DNA, namely adenine, guanine, cytosine and thymidine, base pair mismatches are those base combinations other than the A-T and G-C pairs normally found in DNA. As described herein, a mismatch may be indicated, for example as C/C meaning that a cytosine residue is found opposite another cytosine, as opposed to the proper pairing partner, guanine.

The phrase "DNA insertion or deletion" refers to the presence or absence of "matched" bases

between two strands of DNA such that complementarity is not maintained over the region of inserted or deleted bases.

The term "complementary" refers to two DNA strands that exhibit substantial normal base pairing characteristics. Complementary DNA may contain one or more mismatches, however.

The term "hybridization" refers to the hydrogen bonding that occurs between two complementary DNA strands.

The phrase "flanking nucleic acid sequences" refers to those contiguous nucleic acid sequences that are 5' and 3' to the endonuclease cleavage site.

The term "multiplex analysis" refers to the simultaneous assay of pooled DNA samples according to the above described methods.

The term "substantially pure" refers to a preparation comprising at least 50-60% by weight of the material of interest. More preferably, the preparation comprises at least 75% by weight, and most preferably 90-99% by weight of the material of interest. Purity is measured by methods appropriate for the material being purified, which in the case of protein includes chromatographic methods, agarose or polyacrylamide gel electrophoresis, HPLC analysis and the like.

C>T indicates the substitution of a cytosine residue for a thymidine residue giving rise to a mismatch. Inappropriate substitution of any base for another giving rise to a mismatch or a polymorphism may be indicated this way.

N, N, N', N'-tetramethyl-6-carboxyrhodamine (TAMRA) is a fluorescent dye used to label DNA molecular weight standards which are in turn utilized as an internal standard for DNA analyzed by automated DNA sequencing.

Primers may be labeled fluorescently with 6-carboxyfluorescein (6-FAM). Alternatively primers may be labeled with 4, 7, 2', 7'-Tetrachloro-6-carboxyfluorescein (TET). Other alternative DNA labeling methods are known in the art and are contemplated to be within the scope of the invention.

CEL I has been purified so as to be substantially homogeneous, thus, peptide sequencing of the amino terminus is envisioned to provide the corresponding specific oligonucleotide probes to facilitate cloning of the enzyme from celery. Following cloning and sequencing of the gene, it may be expressed in any number of recombinant DNA systems. This procedure is well known to those skilled in the art and is contemplated to be within the scope of the present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the results of sodium dodecyl sulfate (SDS) polyacrylamide gel analysis of the purified enzyme, CEL I. The positions of molecular weight markers are shown on the side. T indicates the top of the resolving gel.

Figure 2 depicts certain heteroduplex DNA substrates used in performing nucleic acid analyses in accordance with the present invention. Figure 2A depicts a 64-mer which can be terminally labeled at either the 5'-P or the 3'-OH. The nucleotide positions used as a reference in this analysis are indicated irrespective of the number of nucleotide insertions at X in the top strand. The inserted sequences and substrate numbers are indicated in the table. Figure 2B illustrates mismatched basepair substrates used in this analysis, with the identities of nucleotides Y and Z varied as in the accompanying

table to produce various mispaired substrates.

Figure 3 is an autoradiogram demonstrating the effect of temperature on CEL I incisions in different substrates.

Figure 4 is an autoradiogram illustrating the relative incision preferences of CEL I at DNA loops of one nucleotide. Figure 4A shows that in addition to the X=G, the X=C also allows two alternate basepairing conformations. Figure 4B demonstrates that the bottom strand of the substrate is competent for CEL I incision as in the C/C mismatch, #10, in lane 16.

Figure 5 is an autoradiogram of denaturing 15% polyacrylamide gels showing the AmpliTaq DNA polymerase mediated stimulation of purified CEL I incision at DNA mismatches of a single extrahelical nucleotide. F indicates the full length substrate, 64 nucleotides long, labeled at the 5' terminus (*) of the top strand. In panels 5A, 5B and 5C, substrates were treated with varying quantities of CEL I in the presence or absence of DNA polymerase.

Figure 6 is an autoradiogram showing the pH optimum of CEL I incision at the extrahelical G residue in the presence or absence of AmpliTaq DNA polymerase. The top panel shows the CEL I activity in the absence of AmpliTaq DNA polymerase. The bottom panel shows CEL I activity in the presence of polymerase.

Figure 7 is an autoradiogram demonstrating the recognition of base substitution mismatches by purified CEL I in the presence of AmpliTaq DNA

polymerase. (I) indicates the primary incision site at the phosphodiester bond 3' of a mismatched nucleotide. Panel 7A illustrates cleavage of the substrate in the presence of both CEL I and DNA polymerase. In panel 7B, CEL I was omitted.

Figure 8 is an autoradiogram illustrating the ability of CEL I to recognize mutations in pooled DNA samples in the presence of excess wild-type DNA. Lanes 3, 5, 6, 10, 11, 12, and 13 contain single samples containing wild type heteroduplexes. Lanes 4 and 6 contain an AG deletion. Lanes 8 and 9 contain a substrate with an 11 base-pair loop. The samples described above were pooled and treated with CEL I. The results of this "multiplex analysis" are shown in Lane 14.

Figure 9 is an autoradiogram further illustrating the ability of CEL I to recognize mutations in the presence of excess wild-type DNA. 1, 2, 3, 4, 10 or 30 heteroduplexed, radiolabeled PCR products (amplified from exon 2 of the *BRCA1* gene) were exposed to CEL I in a single reaction tube and the products run on a 6% polyacrylamide gel. Lanes 1 and 2 are negative controls run in the absence of CEL I. Lane 3 to 11 contain 1 sample with the AG deletion in the presence of increasing amounts of wild-type non-mutated heteroduplexes.

Figure 10 shows a schematic representative diagram of the *BRCA1* gene and the exon boundaries in the gene.

Figure 11 is a histogram of a sample showing the localization of a 5 base deletion in the 11D exon of *BRCA1* following PCR amplification and treatment

with CEL I. A spike indicates a DNA fragment of a specific size generated by cleavage by CEL I at the site of a mismatch. Panel A shows the results obtained with a 6-FAM labeled primer annealed at nucleotide 3177 of *BRCA1*. Panel B shows the results obtained with a TET labeled primer annealed 73 bases into the intron between exon 11 and exon 12. Panel C represents the TAMRA internal lane size standard. Note that the position of the mutation can be assessed on both strands of DNA.

Figure 12 is a histogram of a sample showing the localization of nonsense mutation, A>T, at position 2154 and a polymorphism C>T at nucleotide 2201 in the 11C exon of *BRCA1* following PCR amplification and treatment with CEL I. Panel A shows a spike at base #700 and Panel B shows a spike at #305 corresponding to the site of the nonsense mutation. Panel C is the TAMRA internal lane standard.

Figure 13 shows the results obtained from four different samples analyzed for the presence of mutations in exon 11A using the methods of the instant invention. Results from the 6-FAM samples are shown. Panel A shows a polymorphism T>C at nucleotide 2430 and a second spike at position #483 corresponding to the site of another polymorphism C>T at nucleotide 2731. Panel B shows only the second polymorphism described in panel A. Panel C shows no polymorphism or mutation. Panel D shows the two polymorphisms seen in panel A.

Figure 14 depicts a gel showing the purification scheme for ARA I mismatch endonuclease of *Arabidopsis thaliana*. Lane 1: Crude extract of cells broken by French Press; Lane 2: 25% - 85% saturated

ammonium sulfate fractionation; Lane 3: Con A-Sepharose affinity column, ARA I was eluted by α-methyl mannoside; Lane 4: Phosphocellulose P-11 column ARA I peak; Lane 5: DEAE Sephacel anion exchange column ARA I peak. The molecular weight standards are shown in lanes indicated with "S".

Figure 15 shows an autoradiogram of a denaturing DNA sequencing gel analysis demonstrating that ARA I cuts mismatched substrates throughout the purification scheme. Lane numbers correspond to those of the purification steps in Figure 14. Panels A, B, C illustrate the ARA I cutting of substrate #2, substrate #4 and substrate #18 (no-mismatch control substrate), respectively. F = full length, I = ARA I cut.

Figure 16 is a schematic diagram of the ARA I based mismatch detection assay.

Figure 17 is an illustration of data obtained from GeneScan analysis of endonucleolytic activity of ARA I on a heteroduplex containing a mismatch.

Figure 18 shows a comparison of the GeneScan mutation detection of ARA I versus CEL I, involving a series of control reactions using the wild type allele of exon 19 of the *BRCA1* gene. This fragment of DNA does not contain any mutations and accordingly no mismatch nicking was observed. Panels A and B show the two strands treated with 7 ng of CEL I and stimulated in mismatch cutting by Amplitaq DNA polymerase. B = (6-FAM); G = (TET). Panels C and D are the two strands treated with 20 ng of purified ARA I without stimulation by Amplitaq DNA polymerase.

Panels E and F show the two strands treated with 20 ng of ARA I stimulated for mismatch cutting by the presence of Amplitaq DNA polymerase.

Figure 19 depicts a side-by-side GeneScan analysis of CEL I and ARA I mismatch detection activity in exon 19 of the *BRCA1* gene. Panels A and B show mismatch cutting using 7 ng of CEL I in the presence of 0.5 units of Amplitaq DNA polymerase. Panels C and D show the cutting of an A nucleotide deletion mismatch by 20 ng of ARA I without Amplitaq DNA polymerase. Panels E and F show the cutting of the same substrate by 2 ng of ARA I stimulated in mismatch cutting by the presence of 0.5 ng units of Amplitaq DNA polymerase. All mutations and polymorphisms detected were confirmed by automated sequencing. These results suggest that ARA I, like the CEL I mutation detection method, can identify mutations that are difficult to detect with SSCP or DNA sequencing.

Figure 20 shows a comparison of the GeneScan mutation detection of ARA I versus CEL I, involving a series of control reactions employing the wild type allele of exon 2 of the *BRCA1* gene. As in figure 18, this gene segment does not contain any mutations, thus no mismatch nicking is observed. Panel A and B show the two strands treated with 7 ng of CEL I stimulated in mismatch cutting by 0.5 units of Amplitaq DNA polymerase. Panels C and D show the two strands treated with 20 ng of purified ARA I without stimulation by Amplitaq DNA polymerase. Panels E and F show the two strands treated with 20 ng of ARA I stimulated for mismatch cutting by the presence of Amplitaq DNA polymerase. Panels G and H show the two strands treated with 2 ng of ARA I stimulated for

mismatch cutting by the presence of 0.5 units of Amplitaq DNA polymerase.

Figure 21 depicts GeneScan analysis of CEL I and ARA I mismatch detection in Exon 2 of the BRCA1 gene. Panels A and B show an AG-deletion mismatch cutting by 7 ng of CEL I in the presence of 0.5 units of Amplitaq DNA polymerase. Panels C and D show the cutting of the AG nucleotide deletion mismatch by 20 ng of ARA I without Amplitaq DNA polymerase. Panels E and F show the cutting of the same substrate by 20 ng of ARA I stimulated in mismatch cutting by the presence of 0.5 ng units of Amplitaq DNA polymerase. Panels G and H show the cutting of the same substrate by 2 ng of ARA I stimulated in mismatch cutting by the presence of 0.5 units of Amplitaq DNA polymerase.

Figure 22 is an autoradiogram demonstrating that mismatch endonuclease activity similar to that of ARA I and CEL I is present in the extracts of 10 other plants.

Figure 23 is an autoradiogram showing that mismatch endonuclease activity similar to ARA I and CEL I is present in the extracts of 11 other plants.

DETAILED DESCRIPTION OF THE INVENTION

The enzymatic basis for the maintenance of correct base sequences during DNA replication has been extensively studied in *E. coli*. This organism has evolved a mismatch repair pathway that corrects a variety of DNA basepair mismatches in hemimethylated DNA as well as insertions/deletions up to four nucleotides long. Cells deficient in this pathway mutate more frequently, hence the genes are called MutS, MutL and MutH etc. MutS protein binds to the

mismatch and MutH is the endonuclease that incises the DNA at a GATC site on the strand in which the A residue is not methylated. MutL forms a complex with MutH and MutS during repair. Homologs of MutS and MutL, but not MutH exist in many systems. In yeast MSH2 (MutS homolog) can bind to a mismatch by itself, but a complex of two MutL homologs (MLH and PMS1) plus a MSH2 has been observed. The human homolog hMSH2 has evolved to bind to larger DNA insertions up to 14 nucleotides in length, which frequently arise by mechanisms such as misalignment at the microsatelite repeats in humans. A role for hMLH1 in loop repair is unclear. Mutations in any one of these human homologs were shown to be responsible for the hereditary form of non-polyposis colon cancer (27, 28).

Celery contains over 40 μg of psoralen, a photoreactive intercalator, per gram of tissue (3). As a necessity, celery may possess a high capability for the repair of lesions of insertion, deletion, and other psoralen photoadducts. Single-strandedness at the site of the lesion is common to base substitution and DNA loop lesions. The data in the following examples demonstrate that celery, *Arabidopsis thaliana* and other plant species possess ample mismatch-specific endonucleases to deal with these potentially mutagenic events.

It has been found that the incision at a mismatch site by CEL I is greatly stimulated by the presence of a DNA polymerase. For a DNA loop containing a single nucleotide insertion, CEL I substrate preference is A $\geq$ G > T > C. For base-substitution mismatched basepairs, CEL I preference is C/C $\geq$ C/A ~ C/T $\geq$ G/G > A/C ~ A/A ~ T/C > T/G ~ G/T ~ G/A ~ A/G > T/T. CEL I shows a broad pH optimum from pH 6 to pH 9. To a lesser extent compared with loop incisions, CEL I is also a single-stranded DNase, and a weak exonuclease.

CEL I possesses novel biochemical activities when compared to other nucleases. Mung Bean Nuclease is a 39 kd nuclease that is a single-stranded DNase and RNase, and has the ability to nick DNA at destabilized regions and DNA loops (19-22). However, it has a pH optimum at 5.0. It is not known whether Mung Bean Nuclease activity can be stimulated by a DNA polymerase as in the case of CEL I. Thus CEL I and Mung Bean Nuclease appear to be different enzymes; however this has not yet been conclusively confirmed.

The mechanism responsible for the AmpliTaq DNA polymerase stimulation of the CEL I activity is presently unknown. One possibility is that the DNA polymerase has a high affinity for the 3'-OH group produced by the CEL I incision at the mismatch and displaces CEL I simply by competition for the site. CEL I may have different affinities for the 3'-OH termini generated by incisions at different mismatches, thereby attenuating the extent that AmpliTaq DNA polymerase can stimulate its activity. The use of a DNA polymerase to displace a repair endonuclease in DNA repair was also observed for the UvrABC endonuclease mechanism (25). It was shown that the UvrABC endonuclease does not turnover unless it is in the presence of DNA polymerase I. The protein factors in vivo that can stimulate the CEL I activity may not be limited to DNA polymerases. It is possible that DNA helicases, DNA ligases, 3'-5' exonucleases or proteins that bind to DNA termini may perform that function.

It is important to note that a 5'-labeled substrate can be used to show a CEL I incision band in a denaturing polyacrylamide gel. Recently, a putative human all-type mismatch incision activity (24) was shown to be related to the human topoisomerase I. This enzyme is unable to release itself from a 5'-

labeled substrate after mismatch nicking due to the formation of a covalent enzyme-DNA intermediate with the 3′ terminus of the DNA nick (26). This covalent protein-DNA complex cannot migrate into the denaturing polyacrylamide gel to form a band. CEL I mismatch nicking has been demonstrated with 5′ labeled substrates. Therefore, CEL I is not a plant equivalent of the topoisomerase I-like human all-type mismatch repair activity.

CEL I appears to be a mannopyranosyl glycoprotein as judged by its tight binding to Concanavalin A-Sepharose resin and by the staining of CEL I with the Periodic acid-Schiff glycoprotein stain. Insofar as is known, no repair enzyme has been demonstrated to be a glycoprotein. Glycoproteins are often found to be excreted from the cell, on cellular membranes or secreted into organelles. However, glycoproteins have also been shown to exist in the nucleus for important functions. The level of a 100 kDa stress glycoprotein was found to increase in the nucleus when Gerbil fibroma cells are subjected to heat shock treatment (27). Transcription factors for RNA polymerase II in human cells are known to be modified with N-acetylglucosamine residues (28, 29). Recently, lactoferrin, an iron-binding glycoprotein, was found to bind to DNA in the nucleus of human cells and it activated transcription in a sequence-specific manner (30). The nuclei of cells infected with some viruses are known to contain viral glycoproteins (31-33). These examples where glycoproteins are known to exist inside the nucleus, not merely on the nuclear membrane or at the nuclear pores, tend to show that glycosylated proteins may be important in the nucleus. CEL I appears to be an example of a glycoprotein that can participate in DNA repair.

The properties of the celery mismatch

endonuclease CEL I resemble those of single-stranded nucleases. The best-suited substrates for CEL I are DNA loops and base-substitution mismatches such as the C/C mismatch. In contrast, loops greater than 4 nucleotides and the C/C mismatch are the substrates worst-suited for the *E. coli mutHLS* mismatch repair system (1,2). Thus CEL I is an enzyme that possesses novel mismatch endonuclease activity.

The following examples are provided to describe the invention in further detail. These examples, which set forth the best mode presently contemplated for carrying out the invention, are intended to illustrate and not to limit the invention.

Example I

Purification of CEL I

Two different CEL I preparations were made up as described below. Their properties are similar except that the less pure preparation (Mono Q fraction) may contain protein factors that can stimulate the CEL I activity.

(i) Preparation of CEL I Mono Q fraction

100 gm of celery stalk was homogenized in a Waring blender with 100 ml of a buffer of 0.1 M Tris-HCl pH 7.0 with 10 μM phenylmethanesulfonyl fluoride (PMSF) (Buffer A) at 4 °C for 2 minutes. The mixture was cleared by centrifugation, and the supernatant was stored at -70°C. The extract was fractionated by anion exchange chromatography on a FPLC Mono Q HR5/10 column. The bound CEL I nuclease activity was eluted with a linear gradient of salt at about 0.15 M KCl.

(ii) Preparation of highly purified CEL I

7 Kg of celery at 4 °C was extracted with a juicer and adjusted with 10X Buffer A to give a final concentration of 1X Buffer A. The extract was concentrated with a 25% to 85% saturation ammonium sulfate precipitation step. The final pellet was dissolved in 250 ml of Buffer A and dialyzed against 0.5 M KCl in Buffer A. The solution was incubated with 10 ml of Concanavalin A-Sepharose resin (Sigma) overnight at 4 °C. The slurry was packed into a 2.5 cm diameter column and washed with 0.5 M KCl in Buffer A. The bound CEL I was eluted with 60 ml of 0.3 M α-D mannose, 0.5 M KCl in Buffer A at 65 °C. The CEL I was dialyzed against a solution of 25 mM KPO_4, 10 μM PMSF, pH 7.4 (Buffer B), and applied to a phosphocellulose column that had been equilibrated in the Buffer B. The bound enzyme was eluted with a linear gradient of KCl in Buffer B. The peak of CEL I activity from this column was further fractionated by size on a Superose 12 FPLC column in 0.2 M KCl, 1 mM $ZnCl_2$, 10 μM PMSF, 50 mM Tris-HCl pH 7.8. The center of the CEL I peak from this gel filtration step was used as the purified CEL I in this study. A protein band of about 34,000 daltons is visible when 5 micrograms of CEL I of the Superose 12 fraction was visualized with Coomassie Blue staining or carbohydrate staining (Periodic acid-Schiff base mediated staining kit, SIGMA Chemicals (5)) on a 15% polyacrylamide SDS PAGE gel as shown in Figure 1. A second band of approximately 36,000 daltons was also visible in the gel. Both bands were stained with the glycoprotein specific stain. The subtle mobility differences observed in the two bands may be due to differential glycosylation. Alternatively, there may be a contaminant in the preparation which co-purifies with CEL I.

Protein determination

Protein concentrations of the samples were determined by the Bicinchoninic acid protein assay (4, Pierce).

Following purification of CEL I enzyme, mutational analysis on experimental and clinical DNA substrates were performed in a suitable gel system. CEL I recognized and cleaved DNA at a variety of mismatches, deletions and insertions. The following examples describe in greater detail the manner in which mutational analysis is practiced according to this invention.

EXAMPLE II

Preparation of heteroduplexes containing various mismatches

DNA heteroduplex substrates of 64 basepairs long were constructed containing mismatched basepairs or DNA loops which were prepared using similar methods reported in Jones and Yeung (34). The DNA loops are composed of different nucleotides and various loop sizes as illustrated in Fig. 2. The DNA duplexes were labeled at one of the four termini so that DNA endonuclease incisions at the mispaired nucleotides could be identified as a truncated DNA band on a denaturing DNA sequencing gel. The oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer and purified by using a denaturing PAGE gel in the presence of 7M urea at 50 °C. The purified single-stranded oligonucleotides were hybridized with appropriate opposite strands. The DNA duplex, containing mismatches or not, was purified by using a nondenaturing PAGE gel. DNA was eluted from the gel slice by using electro-elution in a Centricon unit in an AMICON model 57005 electroeluter. The upper reservoir of this unit has been redesigned to include

water-tight partitions that prevent cross-contamination.

EXAMPLE III

Mismatch endonuclease assay

Fifty to 100 fmol of 5' [^{32}P]-labeled substrate described in Example II were incubated with the Mono Q CEL I preparation in 20 mM Tris-HCl pH 7.4, 25 mM KCl, 10 mM $MgCl_2$ for 30 minutes at temperatures of 0 °C to 80 °C. From one half to 2.5 units of AmpliTaq DNA polymerase was added to the nuclease assay reaction. Ten μM dNTP was included in the reaction mixture where indicated (Figures 2 & 5). The 20 μL reaction was terminated by adding 10 μL of 1.5 % SDS, 47 mM EDTA, and 75% formamide plus tracking dyes and analyzed on a denaturing 15% PAGE gel in 7M urea at 50°C. An autoradiogram was used to visualize the radioactive bands. Chemical DNA sequencing ladders were included as size markers. Incision sites were accurately determined by co-electrophoresis of the incision band and the DNA sequencing ladder in the same lane.

Example IV

The Effect of Temperature on CEL I Incision Activity at single-nucleotide DNA loop and nucleotide substitutions

The CEL I fraction eluted from the Mono Q chromatography of the celery extract was found to specifically nick DNA heteroduplexes containing DNA loops with a single extrahelical guanine (substrate #2) or thymine residue (#3), but not the perfectly basepaired DNA duplex #1 as shown in Fig. 3. In these experiments fifty fmol of heteroduplex #2 (lanes 3-9), #3 (lanes 10-16), perfectly basepaired duplex #1 (lanes 17-23) and single-stranded DNA substrate (lanes

24-30), each labeled at the 5'-terminus with γ-[^{32}p] ATP and T4 polynucleotide kinase at about 6000 Ci/mmol, were incubated with 0.5 μL (10 μg) of the Mono Q fraction of the CEL I preparation in 20 mM Tris-HCl pH 7.4, 25 mM KCl, 10 mM $MgCl_2$ for 30 minutes at various temperatures. Each 20 μL reaction was terminated by adding 10 μL of 1.5% SDS, 47 mM EDTA, and 75% formamide containing xylene cyanol and bromophenol blue. Ten μL of the sample was loaded onto a 15% polyacrylamide, 7 M urea denaturing DNA sequencing gel at about 50 °C, and subjected to electrophoretic separation and autoradiography as previously reported (7). The G+A and the T chemical sequencing reactions were performed as described (7) and used as size markers. CEL I incision produced bands at about 35 nucleotides long. Lines are drawn from the positions of the incision bands to the phosphodiester bonds (I and II) nicked by the endonuclease in the reference sequencing ladder. For a 5'-labeled substrate, when a nuclease nicks 5' of a nucleotide and produces a 3'-OH terminus, the truncated band runs half a nucleotide spacing slower than the band for that nucleotide in the chemical DNA sequencing reaction product lane (34).

Substrate #2 can basepair in two conformations because the inserted G is within a CGCG sequence. Therefore either the G residue in the second or the third nucleotide position can become unpaired, possibly extrahelical in conformation, when this duplex is hybridized:

5'-CGGCG-3' or 5'-CGGCG-3'
3'-G-CGC-5' 5'-GC-GC-5'

Accordingly, two mismatch incision bands were observed, each correlating to the phosphodiester bond

immediately 3' of the unpaired nucleotide. See Fig. 3, lanes 3-9. This slippage can occur in the target sequence only when G or C is in the mismatched top strand. Therefore, the non-paired T residue in substrate #3 gave one incision band at the same relative position as the upper band derived from the substrate #2. See Fig. 3, lanes 10-16. These gel mobilities are consistent with the production of a 3'-OH group on the deoxyribose moiety (6). CEL I increases in activity with temperature up to 45°C as illustrated by the increase in band intensity, see Fig. 3. However, from 65°C to 80°C, specificity is diminished due to DNA duplex denaturation.

EXAMPLE V

Relative Incision Preferences of CEL I

To ascertain whether there is a single endonuclease incision at each DNA duplex, the experiment described in Fig. 3 was repeated with DNA labeled on the 3' terminus of the top strand. If there were only one incision site, initial incision positions revealed by substrates labeled at the 5' or the 3' termini should be at the same phosphodiester bond. In these experiments, substrates were labeled at the 3' termini with [^{32}P] α-dCTP, cold dGTP and the Klenow fragment of DNA polymerase I to about 6000 Ci/mmol. The sample preparation, denaturing gel resolution and autoradiogram analysis are the same as described in Fig. 3 except incubation of 50 fmole of substrate with 10 μg of the CEL I Mono Q fraction was for 30 minutes at a single temperature, 37°C. The DNA sequencing ladders for substrates #4 and #5 are shown in lanes 1-4 to illustrate the DNA sequences used. Lanes 5-8 had no enzyme during the incubation. Lanes 9-12 are mismatch endonuclease incisions of the substrates #2, #4, #5, #3, respectively. A line is

drawn from the position of the incision band to the phosphodiester bond (I) nicked by the endonuclease in the reference sequencing ladder. Lanes 13 and 14 demonstrate the coelectrophoresis of the CEL I incision band with a chemical DNA sequencing ladder to accurately determine the incision position. Relative incision preferences for substrates #2, #3, #4, and #5 are shown in Fig. 4 for the 3' labeled substrates. The mobilities of the incision bands in lanes 9-12 of Fig. 4 indicate that the incision reactions had occurred at the phosphodiester bond immediately 3' of the unpaired nucleotide. Therefore, the incision site is the same for substrates labeled either at the 5' or the 3' terminus. The fact that the DNA incision was found to occur at the same bond position, whether the substrate DNA was labeled at the 5' termini or the 3' termini shows that CEL I is not a DNA glycosylase. A DNA glycosylase mechanism would cause the DNA incision position in the two DNA substrates to be one base apart because a base is excised by the DNA glycosylase.

Precise determination of the incision site was performed as in the example in lane 14 in which the T residue chemical sequencing reaction of the labeled top strand of substrate #2 (lane 13) was mixed with the CEL I incision product of lane 9 and analyzed in the same lane. For a 3'-labeled substrate, when a nuclease nicks 3' of a nucleotide and produces a 5' PO_4 terminus, the truncated band runs with the band for that nucleotide in the chemical DNA sequencing reaction product lane (7). Moreover, the gel mobility, relative to the size standards of chemical DNA sequencing, illustrated that the DNA nick produced a 5'-phosphorylated terminus (6). For a DNA loop with a single nucleotide insertion, the nuclease specificity is A $\geq$ G> T > C. It can be seen in Fig.

4A that a small amount of 5' to 3' exonuclease activity is present in this CEL I preparation.

To test whether CEL I can cut in the bottom strand across from a DNA loop of one nucleotide in the top strand, or whether nicking of the loop-containing strand may lead to secondary CEL I incision across from the nick, the bottom strand that contains no unpaired nucleotides in substrate #2 was labeled at the 3' end and incubated in the presence of CEL I. The extrahelical nucleotide in the top strand, or the DNA nick made by CEL I in the top strand of substrate #2, seen in lane 9 of Fig. 4, did not lead to significant nicking of the bottom strand (lane 18). As a control against the possibility that DNA sequence effect may favor CEL I incision in the top strand and not the bottom strand, CEL I was tested for incision of the bottom strand in the C/C mismatch substrate in lanes 15 and 16. Mismatch incision was made when CEL I was present in lane 16.

In the characterization of the incision site of a repair endonuclease, it is important to determine whether one or two incisions have been made for each lesion. This is normally accomplished by using lesion-containing substrates that have been labeled, in turn, at the four termini of a DNA duplex. This test has been satisfied in the analysis of substrate #2 by using three labeled substrates because of the near absence of incision in the bottom strand. In Fig. 3, lane 4-7 and Fig. 4, lane 9, respectively, the incision of this substrate in both the 5' labeled and the 3' labeled substrates have been compared. The incision site was found to be at the 3' side of the mismatched nucleotide in both cases. The lack of incision on the bottom strand for substrate #2 was demonstrated in lane 18 of Fig. 4. Only the 5' labeled substrate was needed in this case since no

significant bottom strand incision had occurred.

Example VI

Effect of AmpliTaq DNA polymerase on the incisions at DNA loop mismatches

CEL I activity is stimulated by the presence of a DNA polymerase. In Fig. 5, the CEL I incisions at single-nucleotide loop substrates were stimulated by AmpliTaq DNA polymerase to different extents depending on which nucleotides are present in the loop. It was necessary to use different amounts of CEL I to illustrate the AmpliTaq DNA polymerase stimulation. The stimulation of the incision at extrahelical C and extrahelical T substrates are best illustrated in Figs. 5 A & B (compare lanes 4 with lanes 9, and lanes 5 with lanes 10, in the respective panels) where higher CEL I levels are required to show good incision at these mismatches. For extrahelical G and extrahelical A substrates that are among the best substrates for CEL I, AmpliTaq DNA polymerase stimulation can best be illustrated by using a much lower level of CEL I as in Fig. 5. The amounts of AmpliTaq stimulation of CEL I in Fig. 5 were quantified and presented in Table I.

Table I

Quantification of the CEL I incision bands shown in the autoradiogram in Fig. 5.

AmpliTaq	-		+		+/-
Substrate	Panel lane#	Counts	Panel lane#	Counts	
Extrahelical G,band I	A,2	20894	A,7	22101	1.1
Extrahelical A,band I	A,3	19451	A,8	26357	1.4
Extrahelical C,band I	A,4	4867	A,9	12009	2.5
Extrahelical T,band I	A,5	2297	A,10	25230	11.0
Extrahelical G,band I	B,2	12270	B,7	19510	1.6
Extrahelical A,band I	B,3	10936	B,8	24960	2.3
Extrahelical C,band I	B,4	1180	B,9	2597	2.2
Extrahelical T,band I	B,5	700	B,10	21086	30.1
Extrahelical G,band I	C,11	10409	C,13	18649	1.8
Extrahelical G,bandII	C,11	9020	C,13	19912	2.2
Extrahelical A,band I	C,12	7165	C,14	14983	2.1

The Autoradiograms were quantified in two dimensions with an AMBIS densitometer and the amount of signal in each band is given as counts.

Example VII

Optimum pH of CEL I Activity

The pH optimum of CEL I for the extrahelical G substrate was investigated in the absence or presence of the AmpliTaq DNA polymerase. CEL I (9.5 ng) was incubated with 100 fmol of the substrate in a 20 μL reaction in buffers of pH 5-6.5 (imidazole) and pH 7-9.5 (Tris-HCl) for 30 minutes at 37 °C. When used, one half unit of AmpliTaq DNA polymerase was present in the incubation in the top (- polymerase) or bottom panels (+ polymerase), respectively. As shown in Fig. 6, CEL I was found to be active from pH 5.0 to pH 9.5, and showed a broad pH optimum centered about

pH 7.5 (top panel). When AmpliTaq DNA polymerase was present, the incision was stimulated across the whole pH range (bottom panel). The assay method did not use initial kinetics and thus precluded quantitative conclusions on this pH profile of CEL I. However, it is clear that the enzyme works very well in the neutral pH ranges.

Example VIII

Incisions by CEL I at basepair substitutions

Other combinations of mismatched substrates are also recognized by CEL I and incised on one of the two DNA strands of each DNA duplex. Some of these substrates are less efficiently incised compared with those containing DNA loops; therefore 45°C was used for incubation instead of 37 °C. Substrates with the 5′ termini of the top strands labeled were used in this study. The autoradiogram of Fig. 7 shows that mismatches containing a C residue are the preferred mismatch substrates with C/C often better than C/A and C/T. The incisions at these mismatches tend to produce two alternate incision positions, one at the phosphodiester bond 3′ of the mismatched C residue, one at the phosphodiester bond one nucleotide further removed in the 3′ direction. Whether alternate incision sites will be observed for these mismatches within another DNA sequence context has not been investigated. One possible explanation for this phenomenon may be greater basepair destabilization next to a mismatch that contains a C residue than for other base-substitutions. Alternatively, the specific mismatched nucleotide may shift one position to the 3′ side because the next nucleotide is also a C residue and the two residues can exchange their roles in the pairing with the G residue in the opposite DNA strand. For base substitution mismatched basepairs, CEL I

specificity in the presence of AmpliTaq DNA polymerase, with respect to the top strand, is C/C ≥ C/A ~ C/T ≥ G/G > A/C ~ A/A ~ T/C > T/G ~ G/T ~ G/A ~ A/G > T/T (Fig. 7A). Because eubacterial DNA polymerases are known to incise at unusual DNA structures (8), a test was conducted to determine whether AmpliTaq DNA polymerase by itself will incise at the 13 substrates used in Fig. 7. Under extended exposure of the autoradiogram, no mismatch incision by the AmpliTaq DNA polymerase was observed (Fig. 7B).

Example IX

Detection of DNA mutations Using CEL-I and Multiplex Analysis

The sensitivity of CEL I for mismatch detection is illustrated by its ability to detect mutations in pooled DNA samples. DNA was obtained from peripheral blood lymphocytes from individuals undergoing genetic screening at the Fox Chase Cancer Center. Samples were obtained from breast cancer-only, ovarian cancer-only, breast/ovarian cancer syndrome families or from non-breast/ovarian cancer control samples. Unlabeled primers specific for exon 2 of *BRCA1* were utilized to PCR amplify this region of the gene. The wild-type PCR products of exon 2 were labeled with gamma ^{32}P-ATP. Briefly, 10 picomoles of PCR product were purified by the Wizard procedure (Promega). Exon 2 wild-type products were then phosphorylated using T4 kinase and 15 picomoles of gamma ^{32}P-ATP at 6,000 Ci/mmol in 30 μl 1X kinase buffer (70 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 5 mM dithiothreitol) at 37°C for 1 hour. The reactions were stopped with 1 μl 0.5 M EDTA. The reaction volume was brought up to 50 μl with 1X STE buffer (100 mM NaCl, 20 mM Tris-HCl, pH 7.5, 10 mM EDTA) and processed through a Pharmacia Probe Quant column. Labeled DNA (1 pmol/μl in 100 μl) was then used for

hybridization with individual unlabeled PCR amplified experimental samples. For each individual sample, 100 fmol of the unlabeled PCR amplified product was incubated with 200 fmol of the ^{32}P-labeled wild-type PCR product in CEL I reaction buffer (25 mM KCl, 10 mM $MgCl_2$, 20 mM Tris-HCl, pH 7.5). Following denaturation and renaturation, heteroduplexed, radiolabeled PCR products were exposed to CEL I for 30 minutes at 37°C in 1X CEL reaction buffer and stopped via the addition of 10 µl stop mix (75% formamide, 47 mM EDTA, 1.5% SDS, xylene cyanol and bromophenol blue). The heteroduplexes were treated with the enzyme individually (lanes 4-13) or pooled in one sample tube (lane 14) and treated. The products of the reaction were loaded onto a 15% polyacrylamide gel containing 7 M urea and the results are shown in Fig. 8. Out of the 10 samples analyzed, 2 contained an AG deletion (lanes 4 and 7), 2 contained an 11 base-pair loop (lanes 8 and 9), and the other 6 were wild type (lanes 5, 6, 10, 11, 12, and 13). Cleavage by CEL I at the AG deletion resulted in the formation of two bands, one of approximately 151 nucleotides from the top strand, the other at 112 nucleotides from the bottom strand (lanes 4 and 7). Cleavage by CEL I at 11 base-pair loops resulted in the formation of one band at 147 nucleotides from the top strand, and a group of bands at 109 nucleotides in the bottom strand (lanes 8 and 9). Lanes 1, 2 and 3 contain DNA that was not exposed to CEL I as negative controls, lane 15 contains 64 and 34 bp nucleotide markers. As can be seen in lane 14 of the gel, when the samples were pooled and exposed simultaneously to CEL I, the enzyme cleaved at all of the above listed mutations with no loss of specificity. Also, the PCR products of the wild-type samples showed no non-specific DNA nicking.

To further illustrate the ability of CEL-I

to detect mutations in pooled DNA samples, 1, 2, 3, 5, 10 or 30 heteroduplexed, radiolabelled PCR products, (again amplified from exon 2 of the *BRCA1* gene), were exposed to CEL-I in a single reaction tube and the products run on a 6% polyacrylamide gel containing 7M urea. Samples were amplified and radiolabeled as described above. Each pool contained only one sample which had a mutation (AG deletion). The other samples in each pool were wild-type. Lanes 1 and 2 contain control samples which were not exposed to CEL I. In the pooled samples where a mutation was present, CEL-I consistently cleaved the PCR products illustrating the sensitivity of the enzyme in the presence of excess wild-type, non-mutated DNA (Lanes 4, 5, 6, 7, 8, 9, and 11). As a control, heteroduplexed PCR products containing no mutations were analyzed and no cut band corresponding to a mutation appeared (Fig. 9, lanes 3 and 10).

EXAMPLE X

Detection of Mutations and Polymorphisms by CEL-I in Samples Obtained from High Risk Families

PCR primer sets specific for the exons in the *BRCA1* gene have been synthesized at Fox Chase Cancer Center. The gene sequence of *BRCA1* is known. The exon boundaries and corresponding base numbers are shown in table II. Primers to amplify desired sequences can be readily designed by those skilled in the art following the methodology set forth in Current Protocols in Molecular Biology, Ausubel et al., eds, John Wiley and Sons, Inc. (1995). These primers were planned such than in each PCR reaction, one primer is labeled at the 5' termini with a fluorescent-label, 6-FAM, while the other primer is similarly labeled with a label of another color, TET. A PCR product will thus be labeled with two colors such that DNA nicking events in either strand can be observed independently

and the measurements corroborated. A summary of the results is presented in Table III.

TABLE II

EXON BOUNDARIES AND CORRESPONDING BASED NUMBERS IN BRCA1

EXON	BASE #'s
1	1 -100
2	101 - 199
3	200 - 253
5	254 - 331
6	332 - 420
7	421 - 560
8	561 - 665
9	666 - 712
10	713 - 788
11	789 - 4215
11B	789 - 1591
11C	1454 - 2459
11A	2248 - 3290
11D	3177 - 4215
12	4216 - 4302
13	4303 - 4476
14	4477 - 4603
15	4604 - 4794
16	4795 - 5105
17	5106 - 5193
18	5194 - 5273
19	5274 - 5310
20	5311 - 5396
21	5397 - 5451
22	5452 - 5526
23	5527 - 5586
24	5587 - 5711

Fig. 10 depicts a schematic of the exons present in the *BRCA1* gene. Peripheral blood samples from individuals in high risk families were collected and the DNA isolated. The PCR products were amplified using Elongase (BRL) and purified using Wizard PCR Preps (Promega). The DNA was heated to 94°C and slowly cooled in 1X CEL I buffer (20 mM Tris-HCl pH 7.4, 25 mM KCl, 10 mM $MgCl_2$) to form heteroduplexes. The heteroduplexes were incubated in 20 µl 1X CEL I buffer with 0.2 µl of CEL I and 0.5 units of AmpliTaq at 45°C for 30 minutes. The reactions were stopped with 1 mM phenanthroline and incubated for an additional 10 minutes at 45°C. The sample was processed through a Centricep column (Princeton Separations) and dried down. One microliter of ABI loading buffer (25 mM EDTA, pH 8.0, 50 mg/ml Blue dextran), 4 µl deionized formamide and 0.5 µl TAMRA internal lane standard were added to the dried DNA pellet. The sample was heated at 90°C for 2 minutes and then quenched on ice prior to loading. The sample was then loaded onto a 4.25% denaturing 34 cm well-to-read acrylamide gel and analyzed on an ABI 373 Sequencer using GENESCAN 672 software. The 6-FAM labelled primer in this experimental sample was at nucleotide 3177 of the *BRCA1* cDNA (region 11D), the TET labelled primer was 73 nucleotides into the intron between exon 11 and exon 12. Each spike represents the presence of a DNA band produced by the cleavage of the heteroduplex by CEL-I where a mutation or a polymorphism is present. One spike represents the size of the CEL I produced fragment from the 3′ side of the mismatch site to the 5′ 6-FAM label of the top strand. The other spike represents the corresponding fragment in the bottom strand from the 3′ side of the mismatch to the 5′ TET label. The sum of the two

fragments equals one base longer than the length of the PCR product. The 6-FAM panel shows a spike at base #645 from the 6-FAM label and the TET panel shows a spike at base #483 from the TET label, both corresponding to the site of the 5 base deletion at nucleotide 3819 of the *BRCA1* cDNA (Fig. 11).

Analysis of exon 11 in another individual was performed using a 6-FAM-labelled primer at nucleotide 1454 of the *BRCA1* cDNA (Fig. 12). The TET-labelled primer was at nucleotide 2459 (region 11C). The PCR amplified products were amplified and prepared as described above. In this individual, the 6-FAM panel shows a spike at base #700 and the TET panel shows a spike at #305, each spike corresponding to the site of CEL I incision in the respective DNA strand at a nonsense mutation of A>T at nucleotide 2154 of the *BRCA1* cDNA. The 6-FAM panel also shows a spike at base #747 and the TET panel shows a spike at #258 corresponding to the site of a polymorphism C>T at nucleotide 2201 of the *BRCA1* cDNA. The nonsense mutation and polymorphism have been confirmed by sequencing of this particular sample (KO-11) using the ABI 377 Sequencer. Spikes that are marked with an asterisk are also present in the no enzyme control lane and represent PCR product background.

Certain individuals have mutations in another region of exon 11, region 11A, on the schematic in Fig. 10. A 6-FAM-labelled primer at nucleotide 2248 of the *BRCA1* cDNA and a TET labeled primer at nucleotide 3290 were used to amplify this region of exon 11. Following amplification, the samples were processed as described above. The four 6-FAM panels represent CEL-I reactions with 4 different individual samples. The first panel in Fig. 13A, sample #KO-2, shows one spike at #182 corresponding to the site of a polymorphism T>C at

nucleotide 2430 and a second spike at nucleotide #483 corresponding to the site of another polymorphism C>T at nucleotide 2731. The second panel, Fig. 13B, sample #KO-3, shows only the second polymorphism. The third panel, Fig. 13C, sample #KO-7 shows no polymorphism. The fourth panel, Fig. 13D, sample #KO-11, shows two spikes corresponding to the two polymorphisms. It is interesting to note that this sample, KO-11, shows up positive for a nonsense mutation and a polymorphism in the region of exon 11C corresponding to nucleotides 1454-2459 as described above.

TABLE III

SUMMARY OF *BRCA1* MUTATIONS AND POLYMORPHISMS DETECTED BY CEL I

EXON	NUCLEOTIDE POSITION #	TYPE OF MUTATION
2	185	AG deletion
2	188	11 base deletion
11 C	2154	A > T
11 D	3819	5 base deletion
11 D	4168	A > G
11 D	4153	A deletion
11 D	4184	4 base deletion
20	5382	C insertion

EXON	NUCLEOTIDE POSITION #	TYPE OF POLYMORPHISM
11 B	1186	A > G
11 C	2201	T > C
11 A	2430	T > C
11 A	2731	C > T
11 D	3667	A > G

Table IV sets forth the 5' and 3' flanking sequences surrounding the mutations detected by CEL I in the present invention. While not exhaustive, it can be seen from the variability of the flanking sequences surrounding these mutations and polymorphisms that CEL I sensitivity and recognition of mismatched DNA heteroduplexes does not appear to be adversely affected by flanking sequences.

TABLE IV

EFFECT OF FLANKING SEQUENCES ON ENDONUCLEASE ACTIVITY OF CEL I

nucleotide position	EXON	type of change	5' flanking sequence	3' flanking sequence
185	2	AG deletion	5'ATCTT 3' TAGGA	5' AGTGT 3' TCACA
188	2	11bp deletion	5' TTAGA 3' AATCT	5'G the next 4 bp are in intron
1186	11 B	A--> G	5' TAAGC 3' ATTCG	5' GAAAC 3' CTTG
2154	11 C	A--> T	5' GAGCC 3' CTCGG	5' AGAAG 3' TCTTC
2201	11 C	T--> C	5' GACAG 3' CTGTC	5' GATAC 3' CTATG
2430	11 A	T--> C	5' AGTAG 3' TCATC	5' AGTAT 3' TCATA
2731	11 A	C--> T	5' TGCTC 3' ACGAG	5' GTTTT 3' CAAAA
3667	11 D	A--> G	5' CAGAA 3' CTCTT	5' GGAGA 3' CCTCT
3819	11 D	5 bp deletion	5' GTAAA 3' CATTT	5' CAATA 3' GTTAT
4153	11 D	A deletion	5' TGATG 3' ACTAC	5' AGAAA 3' TCTTT
4184	11 D	4 bp deletion	5' AATAA 3' TTATT	5' GAAGA 3' CTTCT
4168	11 D	A--> G	5' AACGG 3' TTGCC	5' CTTGA 3' GAACT
5382	20	C insertion	5' ATCCC 3' TAGGG	5' AGGAC 3' TCCTG

As can be seen from the above described examples, utilization of CEL I has distinct advantages over methods employing other mismatch repair systems during analysis of mutations in the clinical setting. These advantages are summarized in Table V.

TABLE V

Comparison of the advantages of methods employing CEL I over current mismatch detection methods:

	S1 nuclease method (7)	DNA mismatch glycosylases (8)	MutS binding assay (9)	Chemical cleavage method (10)	T4 endo-nuclease VII (11)	RNase nicking mismatched RNA:DNA (12)	Automated DNA sequencing	ddNTP SSCP finger-printing	Plant mismatch endo-nuclease CEL I
Assay at neutral pH	no	yes	yes	yes	yes	yes	yes	yes	yes
Applicable to mutations of unknown positions	yes	no	yes	yes	yes	yes	yes	yes	yes
Applicable to all basepair substitutions	unknown	with difficulty	with difficulty	with difficulty	yes	no	yes	yes	yes
Applicable to DNA loops,	yes	no	with difficulty	multiple bands	yes	unknown	yes	yes	yes
Advantage of single major band in loop detection	no	no	yes	no	yes	no	no	no	yes
Advantage of little influence by sequence specificity	no	unknown	yes	unknown	cuts w/o mismatch	unknown	no	with difficulty	yes
Advantage of no RNA instability	yes	yes	yes	yes	yes	no	yes	yes	yes
Ability to show the position of a detectable mutation	yes	yes	no	yes	yes	yes	yes	with difficulty	yes
Ability to lower background with DNA polymerase and DNA ligase recycling reaction	no	no	no	no	with difficulty	no	no	no	yes
Advantage to multiplex samples of same color	unknown	n	with difficulty	yes	unknown	no	no	no	yes
Advantage to analyze targets of 1 Kbp--3 Kbp	unknown	unknown	with difficulty	yes, up to 1 Kbp	unknown	no	no	no	yes

EXAMPLE XI

As noted above, many plant species synthesize efficient endonuclease enzymes. In accordance with the present invention, a novel endonuclease, ARA I has been isolated from Arabidopsis thaliana. This endonuclease is quite similar to CEL I in many respects. *Arabidopsis thaliana* is considered to provide a model system for studies in plant molecular biology and biochemistry. Advantages of the *Arabidopsis* system include a short life cycle of about 26 days, the small size of the plant, the diploid nature of the genome, and especially, the small size of the genome compared with most higher plants and animals. The *Arabidopsis* genome, at 7×10^7 basepairs, is only about 10 times larger than that of *E. coli* (4×10^6 basepairs), making genetic cloning of the mismatch endonuclease and genetic manipulation substantially easier than in higher plants and humans, both containing about 2×10^9 basepairs. Thus the finding of the mismatch endonuclease ARA I in *Arabidopsis*, and the ability to use ARA I to perform mutation detection, are important steps leading to the the application of these mismatch endonuclease in mutation detection.

Preparation of highly purified ARA I

The purification procedure of ARA I is very similar to that disclosed for CEL I. This is not unexpected as the data indicate that the two enzymes are substantially similar.

Two hundred and fifty grams of callus of *Arabidopsis thaliana* ecotype *Columbia* were grown on minimal salt agar and stored frozen. The callus was resuspended in a Waring blender in Buffer A (0.1M Tris-HCl, pH 8.0 with 10 micromolar phenylmethanesulfonyl fluoride (PMSF)). The suspended

cells were broken by two passages through a French Pressure cell at 20,000 PSIG to produce the crude lysate. The crude lysate was cleared by centrifugation, and the supernatant was adjusted to 25% saturation in ammonium sulfate with solid ammonium sulfate at 4°C. After two hours, the solution was centrifuged, and the supernatant was adjusted to 85% saturation in ammonium sulfate. The solution was again centrifuged and the pellet was dissolved in 160 ml of Buffer A containing 0.5M NaCl. Five ml of Con A resin was added to this solution and the mixture was tumbled overnight at 4° for 3 hours. The slurry was packed into a 2.5 cm diameter column and washed with 0.5M KCl in buffer A. The bound ARA I was eluted with about 60 ml of 0.5 M alpha-methylmannoside, 0.5M NaCl, in Buffer A at 4°C. The eluted ARA I was dialyzed against a solution of Buffer B (25mM KPO4, 10 micromolar PMSF, pH 7.4) and applied to a phosphocellulose column equilibrated in Buffer B. The bound enzyme was eluted with a gradient of KCl in Buffer B. The eluted peak from P-11 was dialyzed against buffer A and concentrated by passage through a column of Mono Q anion exchanger. The ARA I eluted from the Mono Q step was purified several thousandfold, however, the preparation was not yet homogeneous.

The protein composition of the various purification steps was analyzed by a 4% to 20% polyacrylamide gradient SDS gel electrophoresis. In the gel, which is shown in Figure 14, the lanes are as follows: 1. extract preparation, 2. ammonium sulfate precipitation, 3. Con-A Sepharose affinity column chromatography, 4. phosphocellulose P-11 chromatography; and 5. final purification over a DEAE Sephacel anion exchange column give rise to a over a 10,000 fold purification of ARA I. Protein in the gel

was visualized with staining with Coomassie Blue R-250.

PCR products were amplified using Amplitaq (Perkin-Elmer) and purified using Wizard PCR Preps (Promega). The DNA was heated to 94 °C and slowly cooled in 1X ARA I buffer (20 mM Tris-HCl, pH 7.4, 25 mM KCl, 10 mM $MgCl_2$) to form heteroduplexes. The heteroduplexes were incubated in 20 µl 1X ARA I buffer with 0.2 µl ARA I (0.01 µg) and 0.5 units of Amplitaq (Perkin-Elmer) at 45°C for 30 minutes. The reactions were stopped with 1 mM phenanthroline and incubated for an additional 10 minutes at 45 °C. The samples were processed through a Centricep column (Princeton Separations) and dried down. One microliter of ABI loading buffer (25 mM EDTA, pH 8.0, 50 mg/ml Blue dextran) 4 µl deionized formamide and 0.5 µl TAMRA internal lane standard were added to the dried DNA pellet. The sample was heated at 90°C for 2 minutes and then quenched on ice prior to loading. The sample was then loaded onto a 4.25 % denaturing 34 cm well-to-read acrylamide gel and analyzed on an ABI 373 Sequencer using GeneScan 672 Software. The vertical axis of the electropherogram is relative fluorescence units. The horizontal axis of the electropherogram is DNA length in nucleotides.

Throughout purification, the presence of mismatch endonuclease activity catalyzed by ARA I was readily observable. Figure 15 shows an autoradiogram of denaturing DNA sequencing gel analysis of ARA I-cut mismatched substrates. The lanes in the gel correspond to the lanes containing the various stages of purified ARA I shown in Figure 14. In Figure 15, Panels A, B, C illustrate the ARA I cutting of substrate #2 with the extrahelical G nucleotide, substrate #4 with the extrahelical A nucleotide, and substrate #18, a no-mismatch control substrate,

respectively. F refers to full length substrates whereas I refers to ARA I cut substrates which produced a fragment 35 nucleotides long.

Preparation of GeneScan targets

Genescan targets were prepared as described for the CEL I studies.

Peripheral blood samples from individuals at high risk for breast/ovarian cancer were collected and the DNA isolated used as PCR templates. PCR primers specific for the exons in the *BRCA1* gene were synthesized with a 6-FAM dye at the 5' end of the forward primer, and a TET dye at the 5' end of the reverse primer. PCR products for an exon were hybridized to form heteroduplexes and reacted with ARA I. The products were resolved by the ABI 373 Automated DNA Sequencer and analyzed by GeneScan 672 software.

A schematic diagram of the ARA I mismatch detection assay is shown in Figure 16. The PCR products of a wild type BRCA1 allele and a mutant BRCA1 allele (with an AG deletion in this example) were mixed. After denaturation by heat, and reannealing, heteroduplexes were formed such that in some of them, the extra AG bases formed a loop in the top strand. In others, the extra CT bases formed a loop in the bottom strand. The looped strands were color labeled by having a color dye marker at the 5' termini of each PCR primer used to create the DNA fragment. ARA I cuts the loops at the 3' side of the mismatch, similar to the mismatch cutting by CEL I. As a result, a truncated blue (6-FAM) band and a truncated green (TET) band are produced. The lengths of these two bands independently pin point the location of the mutation in the fragment.

An illustration of the data obtained from

GeneScan analysis based ARA I cutting of a heteroduplex containing a mismatch is shown in Figure 17. The resolution of the ARA I treated DNA fragments is done in a denaturing polyacrylamide gel in an automated DNA sequencer Model 373 of Perkin Elmer. In the denaturing polyacrylamide gel, the fastest running fluorescent signals are the residual PCR primers. The two ARA I fragments, one blue (6-FAM) and one green (TET) follow at their respective size positions. The slowest migrating band is the uncut full length PCR product. These bands are displayed by the GeneScan software as peaks in the fluorogram at the bottom of Figure 17. The GeneScan software, using molecular weight standards of red color (TAMRA) run within the same lane, identifies the sizes of the blue band and the green band. The length of the ARA I generated fragments from the two colored ends of the PCR product independently pinpoint the location of the mutation.

Figures 18-21 provide a simultaneous comparison of Genescan mutation detection of ARA I versus CEL I, using the same BRCA1 gene PCR products described in the previous examples. As can be seen from the data, CEL I and ARA I appear to have identical enzymatic activity on the substrates tested. Thus ARA I, like CEL I is an important new endonuclease that can be used to facilitate the identification of mutations in DNA. As such, the enzyme provides a valuable addition to the arsenal of reagents utilized in genetic screening assays.

EXAMPLE XII

In support of the claim that the mismatch endonucleases in *Arabidopsis*, and indeed in most plants, are substantially similar to that claimed for CEL I of celery, the data for the mismatch detection by the extracts of 10 plants, besides *Arabidopsis*, is

presented. The indicated plant extracts, namely, alfalfa, mung bean, cabbage, cauliflower, Chá há, lettuce, parsley, celery-cabbage, tomato and broccoli were made by homogenizing one part plant with one part buffer A in a Waring blender. As shown in Figure 22, one microliter of the crude homogenate was assayed with top strand 5' ^{32}P labeled substrate #4 (extrahelical A substrate). Lanes G+A and T are Maxam and Gilbert DNA sequencing ladders used to determine the exact cut site in the top strand. The mismatch endonuclease cut produced a 35 nucleotide long fragment visible in lanes 1 to 11. The activity was seen in root, shoot, stalk, leaves, flower, and fruit of these plants, illustrating its ubiquitous nature. Lanes 12-22, which correspond to lanes 1-11, serve as negative controls using no-mismatch substrate #1. Chá há is a Vietnamese vegetable that resembles celery stalk and is rich in nucleases.

To further illustrate the ubiquitous nature of mismatch endonuclease activity similar to ARA I and CEL I, extracts of a group of 11 plants (including all of the plants from Figure 22, plus asparagus) were analyzed for enzymatic activity. See Figure 23. The plant extracts identified in Figure 22 were used to cut a mismatch substrate #2, (extrahelical G substrate). One microliter of the crude homogenate was assayed with top strand 5' ^{32}P labeled substrate #2. Lanes G, G+A, C, and T are Maxam and Gilbert DNA sequencing ladders used to determine the exact cut site in the top strand. The mismatch endonuclease cut produced 34 and 35 nucleotide long fragments visible in lanes 1 to 17. Two bands were seen for the mismatch cut because of mismatch slippage at the two consecutive G residues in the mismatched substrate. The activity was seen in root, shoot, stalk, leaves, flower, and fruit of these plants. Lanes 12-22

correspond to lanes 2-11, but with the no-mismatch substrate #1. Lanes 14-17 illustrate that the mismatch endonuclease activities from four of the plants are shown to be mannosyl proteins similar to CEL I and ARA I. These activities were bound to a ConA-Sepharose resin and then eluted with mannose buffer. Lanes 18-21 are controls of lanes 14-17 except substrate #1, with no mismatch, was used. It is clear from Figures 22 and 23 that the CEL I and ARA I mismatch endonucleases are substantially similar and conserved among plants both in terms of mismatch cutting ability, abundance of activity, and the mannosyl nature of the glycoproteins.

The description and examples set forth above relate to preferred embodiments of the invention. Other embodiments may be apparent to those skilled in the art. Therefore, the invention is not limited to the particular embodiments described and exemplified, but may be capable of modification or variation without departing from the spirit of the invention, the full scope of which is delineated by the appended claims.

REFERENCES

1. Modrich, P. (1994) *Science* ***266***, 1959- 1960.

2. Su, S.-S., Lahue, R.S., Au, K.G., and Moldrich, P. (1988) *J. Biol. Chem.* ***263***, 5057-5061.

3. Finkelstein, E., Afek U., Gross, E., Aharoni, N., Rosenberg, L., and Halevy, S. (1994) *International Journal of Dermatology* ***33***, 116-118.

4. Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., and Klenk, D. C. (1985) *Analytical Chemistry* ***150***, 76-85.

5. Gregory D. J., Culp, D. J., and Jahnke, M. R. (1990) *Analytical Biochem.* ***185***, 324-330.

6. Yeung, A.T., Mattes, W.B., Oh, E.Y., and Grossman, L. (1983) *Proc. Nat. Acad Sci. USA*, **80**, 6157-6161.

7. Yeung, A. T., Dinehart, W. J. and Jones, B. K. (1988) *Nucleic Acids Res*, **16**, 4539-4554.

8. Lyamichev, V., Brow, M. A. D., and Dahlberg, J. E. (1993) *Science* **260**, 778-783.

9. Ramotar, D., Auchincloss, A. H., and Fraser, M. J. (1987) *J. Biol. Chem.* **262**, 425-31.

10. Chow, T. Y.-K., and Resnick, M. A. (1987) *J. Biol. Chem.* **262**, 17659-17667.

11. Wyen, N.V., Erdei, S., and Farkas, G.L. (1971) *Biochem Biophys. Acta.* **232**, 472-83.

12. Brown, P. H., and Ho, D. T. (1987) *Eur. J. Biochem.* **168**, 357-364.

13. Hanson, D.M. and Fairley, J. L. (1969) *J. Biol. Chem.* **244**, 2440-2449.

14. Nucleases, eds. Linn, S. M., Lloyd, R. S., and Roberts, R. J. Cold Spring Harbor Laboratory Press, 1993.

15. Holloman, W. K., Rowe, T. C., and Rusche, J. R (1981) *J. Biol. Chem.* **256**, 5087-5094.

16. Badman, R. (1972) *Genetic Res., Camb.* **20**, 213-229.

17. Shank T. E., Rhodes, C. Rigby, P. W. J., and Berg, P. (1975) *Proc. Nat. Acad. Sci. USA*, **72**, 989-993.

18. Maekawa, K., Tsunasawa, S., Dibo, G., and Saklyama, F. (1991) *Eur. J. Biochem.* **200**, 651-661.

19. Kowalski, D., Kroeker, W. D., and Laskowski, M. Sr. (1976) *Biochemistry* **15**, 4457-4462.

20. Kroeker, W. D., Kowalski, D., and Laskowski, M. Sr. (1976) *Biochemistry* **15**, 4463-4467.

21. Ardelt, W., and Laskowski, M., Sr. (1971) *Biochem. Biophys. Res., Commun.* **44**, 1205- 1211.

22. Kowalski, D. (1984) *Nucleic Acids. Res.* **12**, 7071-7086.

23. Strickland, J. A., Marzilli L. G., Puckett, Jr., J. M., and Doetsch, P. W. (1991) *Biochemistry* **30**, 9749-9756.

24. Doetsch, P. W., McCray, W. H., Lee, K., Bettler, D. R., and Valenzuela, M. R. L. (1988) *Nucleic Acids Res.* **16**, 6935-6952.

25. Caren, P. R., Kushner, S. R., and Grossman, L, (1985) *Proc. Nat. Acad. Sci. USA* **82**, 4925-4929.

26. Yeh, Y.-C., Liu, H.-F., Ellis, C.A., and Lu, A.-L. (1994) *J. Biol. Chem.* 269, 15498-15504.

27. Welch, W. J., Gerrels, J. I., Thomas, G. P., Lin, J. J.-L., and Feramisco, J. R (1983) *J. Biol. Chem.* **258**, 7102-7111.

28. Jackson, S. P., and Tjian, R (1989) *Proc. Nat. Acad. Sci. USA* **86**, 1781-1785.

29. Jackson, S. P., and Tjian, R (1988) *Cell* **55**, 125-133.

30. He, J., and Furmanski, P. (1995) *Nature* **373**, 721-724.

31. Peeples, M. E. (1988) *Yirology* **162**, 255-259.

32. Buckley, A., and Gould, E. A. (1988) *J. Gen. Virology*, **69**, 1913-1920.

33. Gauffre, A., Viron, A., Barel, M., Hermann, J., Puvion, E., and Frade, R. (1992) *Molecular Immunology* **29**, 1113-1120.

34. Jones, B. K. and Yeung, A. T. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8410-8414.

35. Sarker, et al., (1992) *Nucleic Acids Research* **20**:871-878.

36. Meyers, R.M. et al., (1986) *CSHSQB* **52**:275.

WHAT IS CLAIMED IS:

1. A method for determining a mutation in a target sequence of single stranded polynucleotide with reference to a non-mutated sequence of a polynucleotide that is hybridizable with the polynucleotide including said target sequence, wherein said sequences are amplified, labeled with a detectable marker, hybridized to one another, exposed to endonuclease and analyzed for the presence of said mutation, the improvement comprising the use of a mismatch endonuclease, the activity of said endonuclease comprising:

a) detection of all mismatches between said hybridized sequences;

b) recognition of sequence differences in polynucleotide strands between about 100bp and about 3kb in length; and

c) recognition of said mutation in a target polynucleotide sequence without substantial adverse effect caused by flanking polynucleotide sequences.

2. The method as claimed in claim 1 wherein said endonuclease is derived from celery.

3. The method as claimed in claim 1 wherein said polynucleotide is DNA.

4. The method as claimed in claim 2 wherein the sequences exposed to said endonuclease are also exposed to a protein selected from the group consisting of DNA ligase, DNA polymerase, DNA helicase, 3'-5' DNA Exonuclease, DNA binding proteins that bind to DNA termini or a combination of said proteins, thereby reducing non-specific DNA cleavage.

5. The method as claimed in claim 1 wherein the sequences exposed to said endonuclease are also exposed to DNA polymerase.

6. The method as claimed in claim 2 wherein target DNA is analyzed on a multiplex grid.

7. The method as claimed in claim 2 wherein said polynucleotide is cDNA.

8. The method as claimed in claim 1, wherein said sequences are analyzed on a DNA sequencing gel thereby identifying the location of the mutation in a target DNA strand relative to DNA sequencing molecular weight markers.

9. The method as claimed in claim 1 wherein said mutation is determined as a means of genetic screening for cancer.

10. The method as claimed in claim 1 wherein said mutation is determined as a means of detecting of genetic alterations leading to birth defects.

11. A method for determining a mutation in a target sequence of single stranded polynucleotide with reference to a non-mutated sequence of a polynucleotide that is hybridizable with the polynucleotide including said target sequence, wherein said sequences are amplified, labeled with a detectable marker, hybridized to one another, exposed to endonuclease and analyzed for the presence of said mutation, the improvement comprising the use of a mismatch endonuclease derived from celery, the activity of said endonuclease comprising:

a) detection of all mismatches between said hybridized sequences;

b) recognition of sequence differences in polynucleotide strands between about 100bp and about 3kb in length;

c) recognition of said mutation in a target polynucleotide sequence without substantial adverse effect caused by flanking polynucleotide sequences;

d) recognition of polynucleotide loops and insertions between said hybridized sequences; and

e) recognition of said mutation in a target polynucleotide sequence in which a DNA polymorphism or another mutation is also present.

12. A mismatch endonuclease for determining a mutation in a target sequence of single stranded mammalian polynucleotide with reference to a non-mutated sequence in a polynucleotide that is hybridizable with the polynucleotide including said target sequence, said endonuclease being in substantially pure form and effective to:

a) detect all mismatches between said hybridized sequences;

b) recognize sequence differences in polynucleotide strands about 100 bp and about 3 kb bp in length;

c) recognize said mutations in a target polynucleotide sequence without substantial adverse effect caused by flanking DNA sequences; and

13. The endonuclease of claim 12 wherein said endonuclease recognizes polynucleotide loops and insertions between said hybridized sequences.

14. The endonuclease of claim 12 wherein

said endonuclease recognizes said mutation in a target polynucleotide sequence in which a DNA polymorphism or another mutation is also present.

15. The endonuclease of claim 12 derived from a plant source.

16. The endonuclease of claim 15, wherein said plant source is celery.

17. The endonuclease of claim 15, wherein said plant source is *Arabidopsis thabiana*.

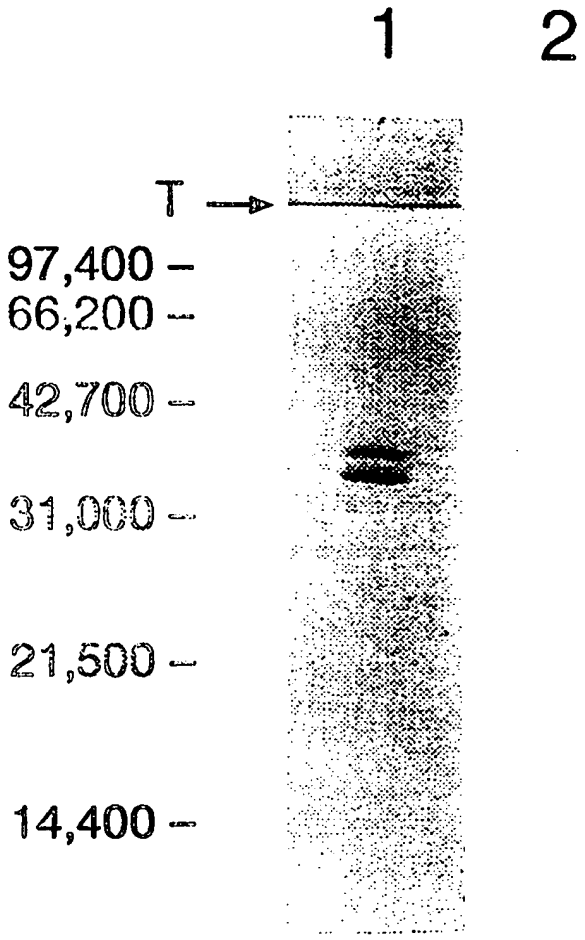
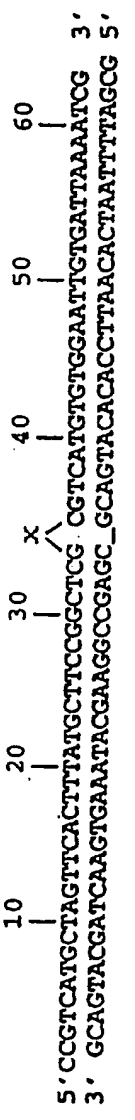


Figure 1

Fig. 2A



Substrate #	1	2	3	4	5
Loop X	-	G	T	A	C

Fig. 2B

Substrate #	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Y	A	A	A	C	C	C	G	G	G	T	T	T	A	C	T
Z	A	C	G	A	C	T	A	G	T	C	G	T	T	G	A

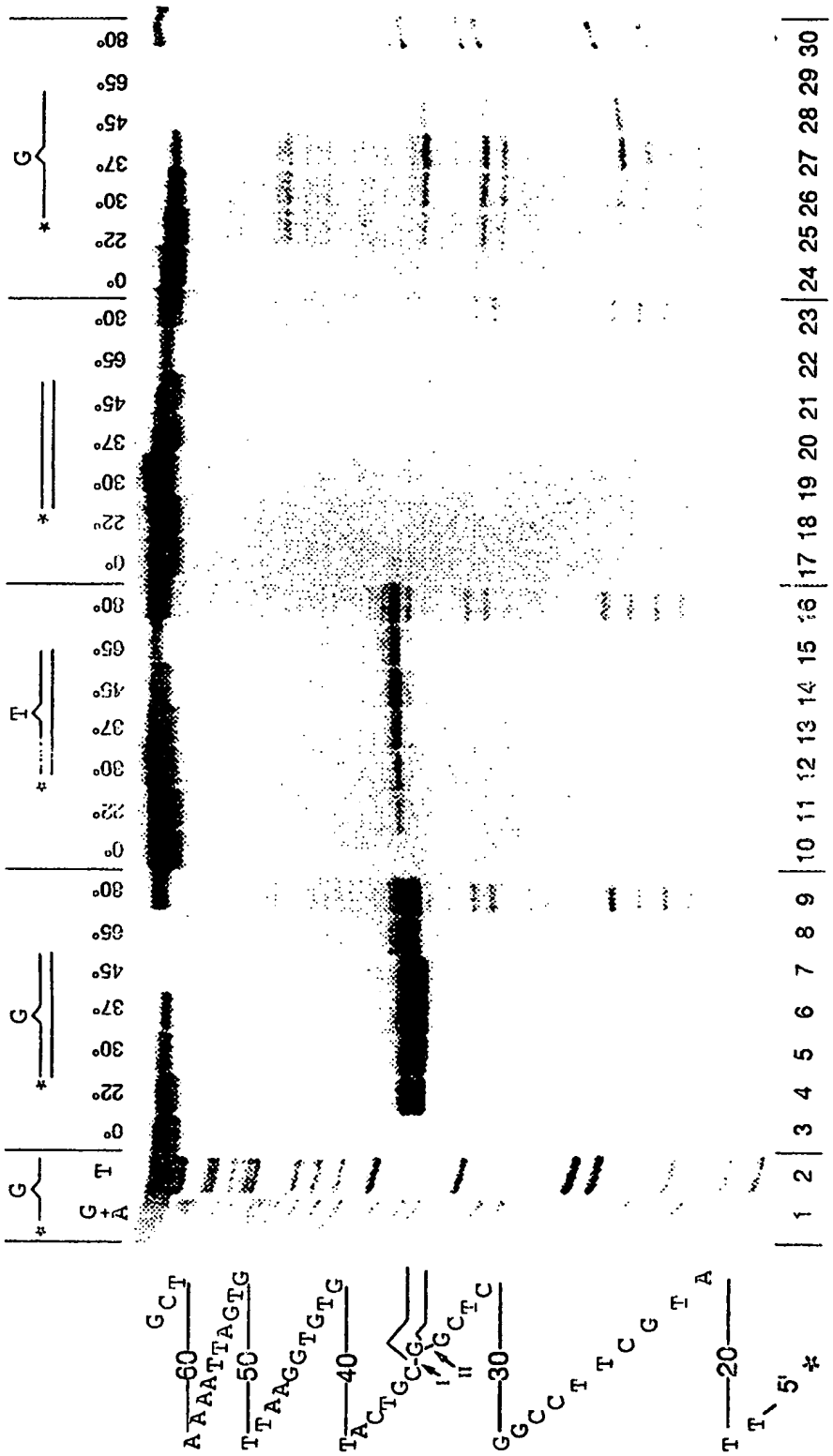


Figure 3

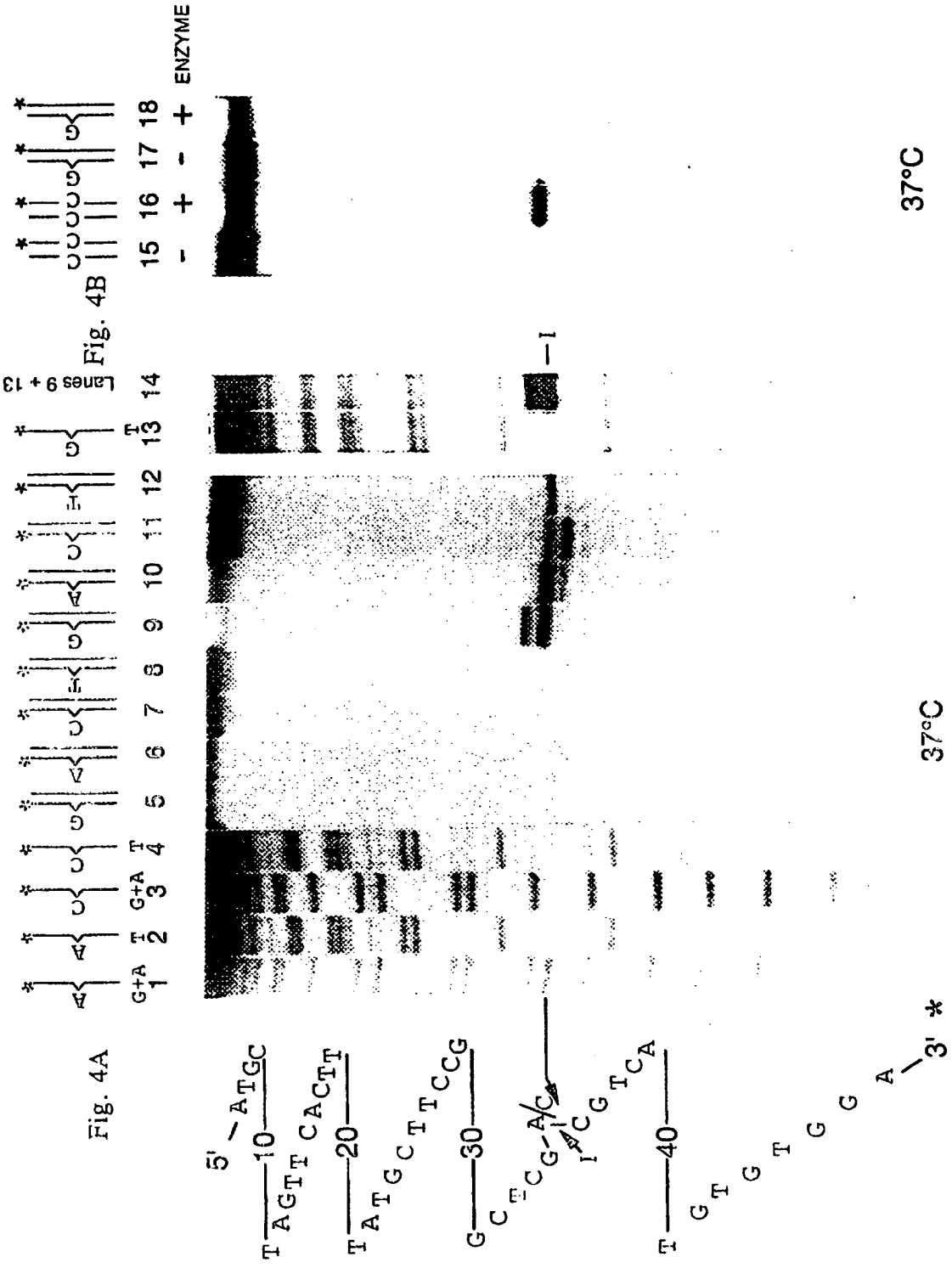


Fig. 4A
Fig. 4B
Lanes 9 + 13
ENZYME
37°C
37°C
5'
3'
10
20
30
40

Fig. 5A

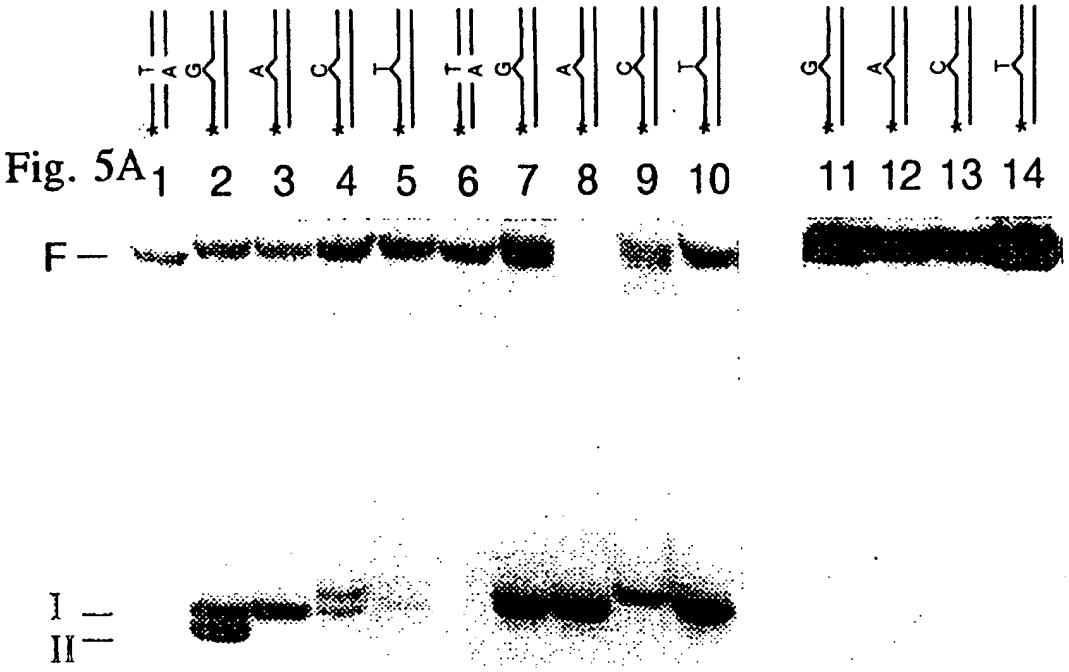
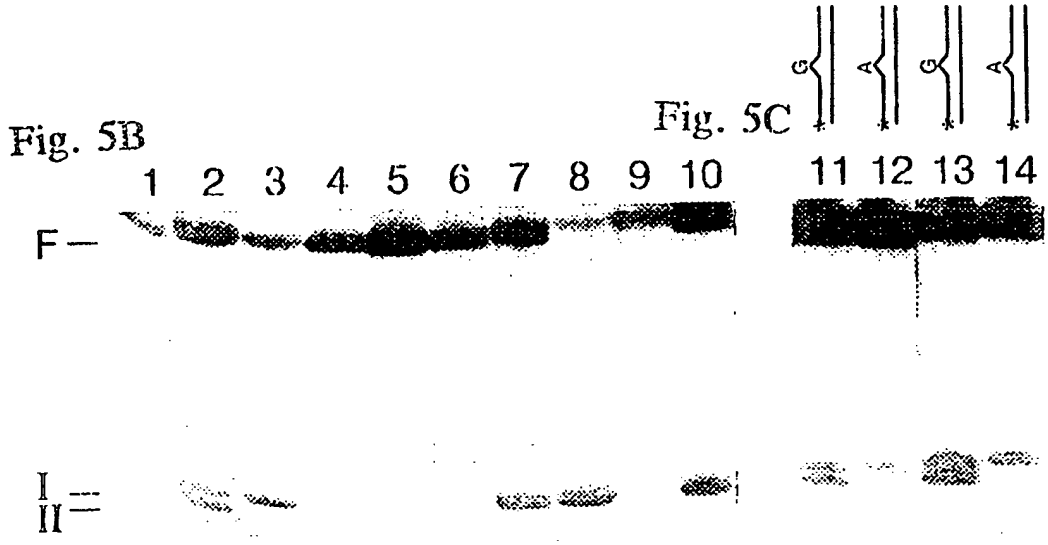


Fig. 5B

Fig. 5C



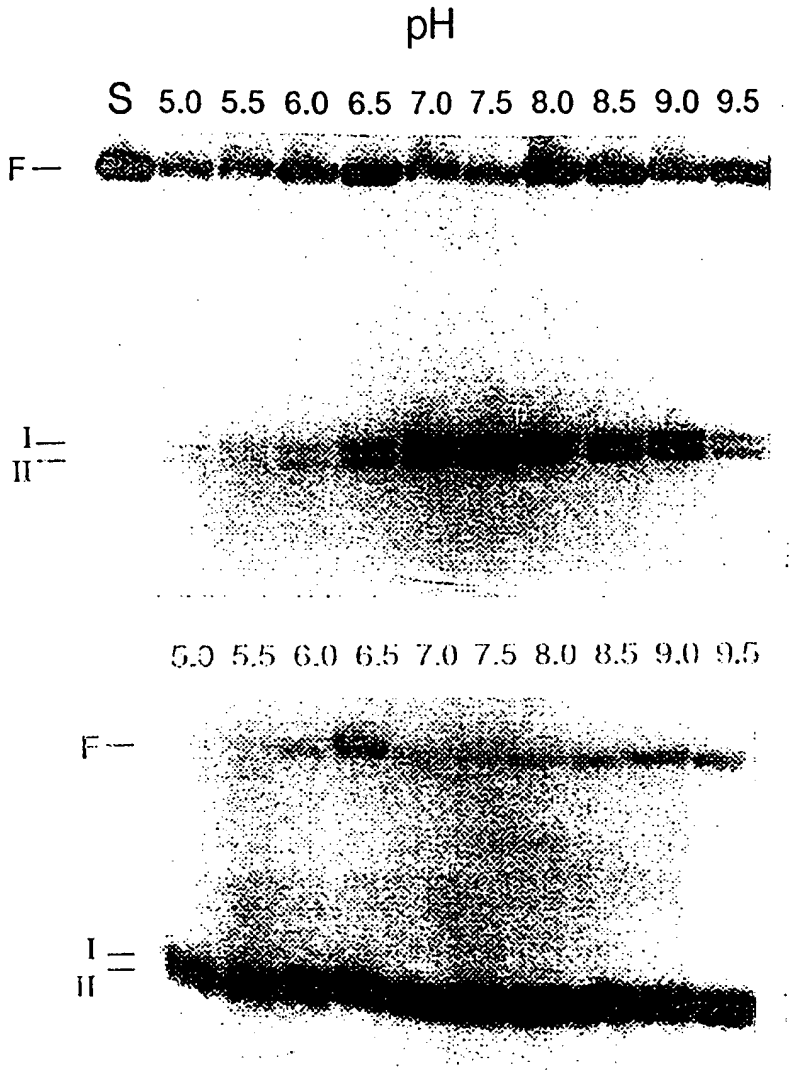

pH
S 5.0 5.5 6.0 6.5 7.0 7.5 8.0 8.5 9.0 9.5
F
I
II
5.0 5.5 6.0 6.5 7.0 7.5 8.0 8.5 9.0 9.5
F
I
II

Figure 6

Fig. 7A

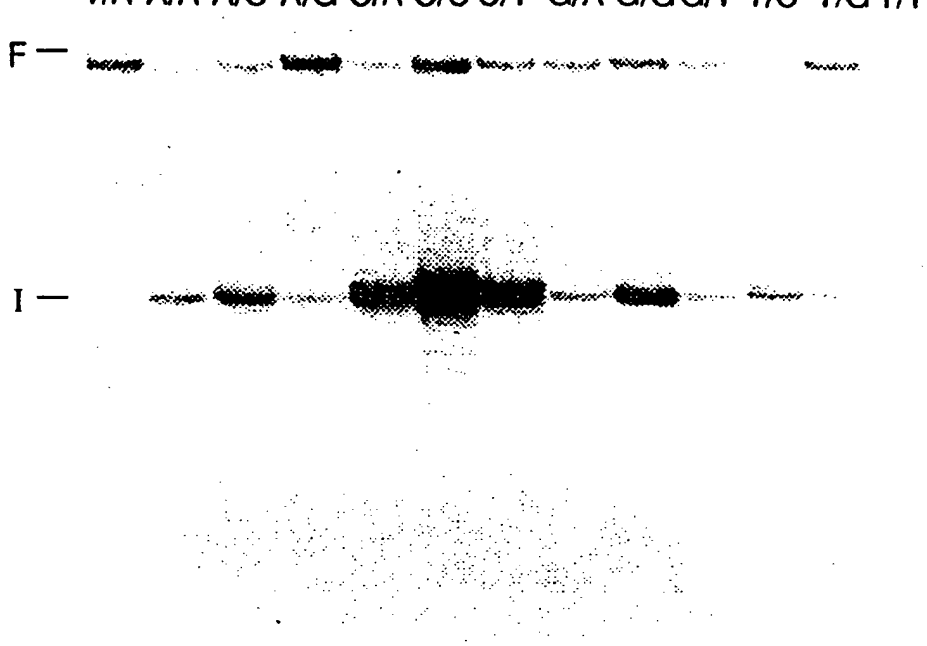
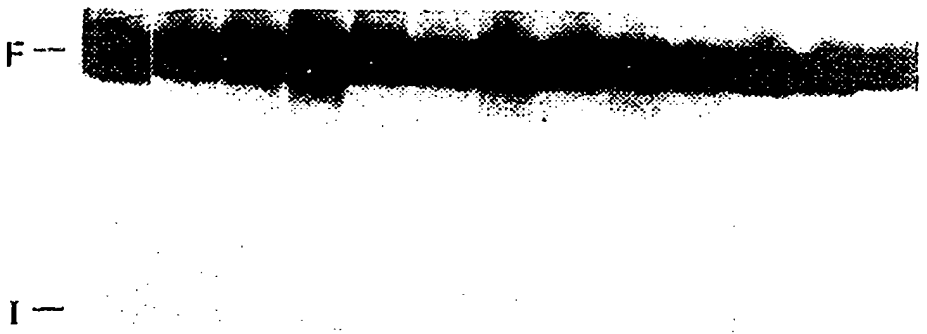


Fig. 7B



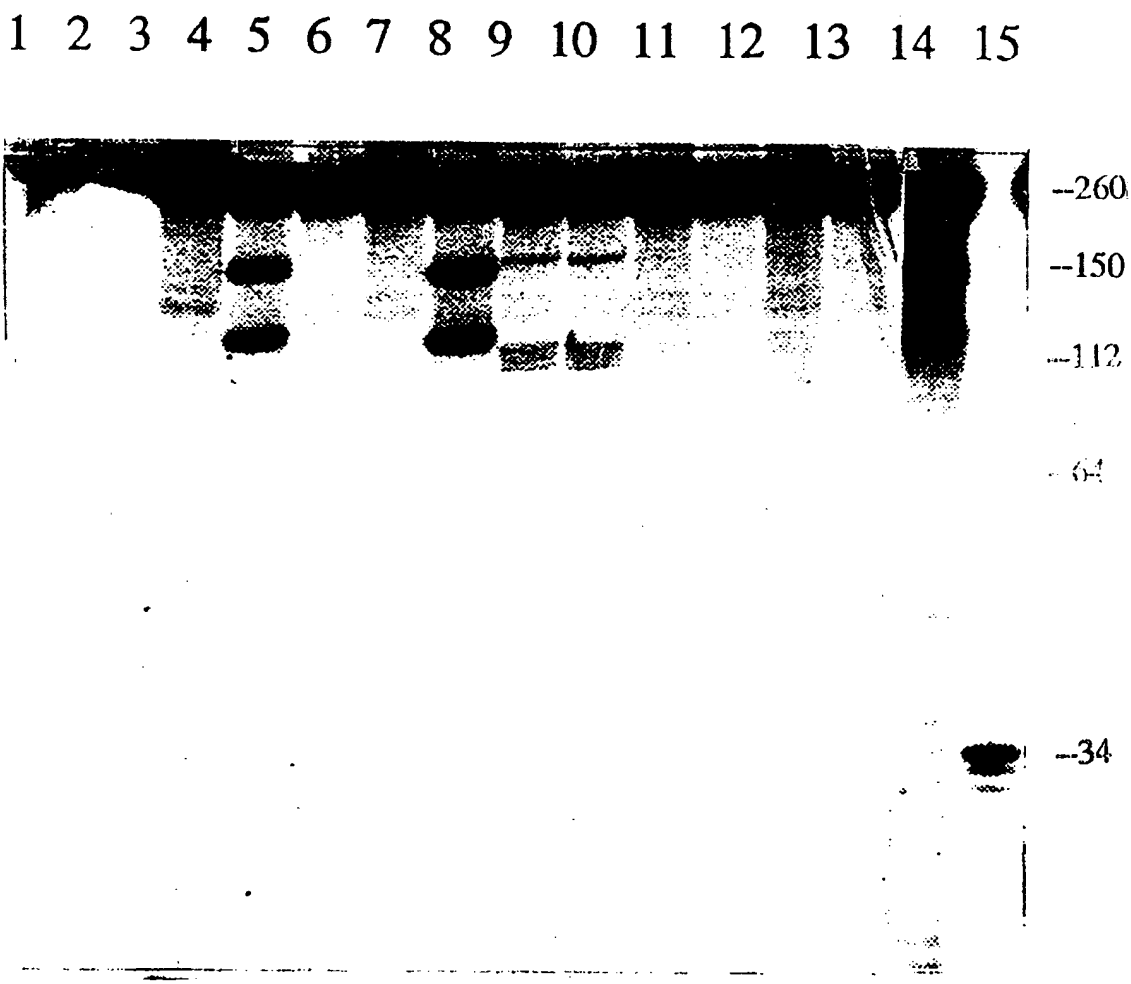


Figure 8

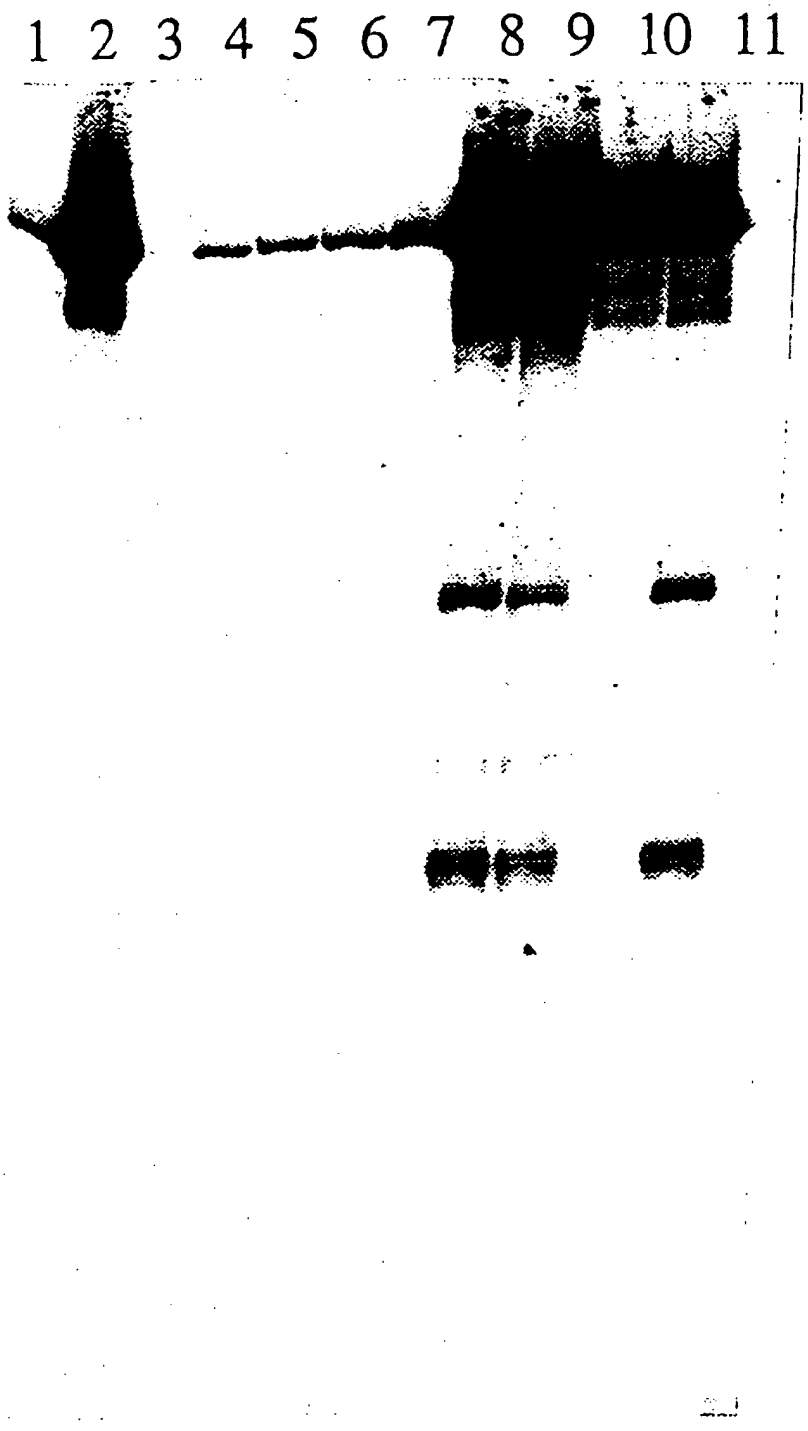


Figure 9

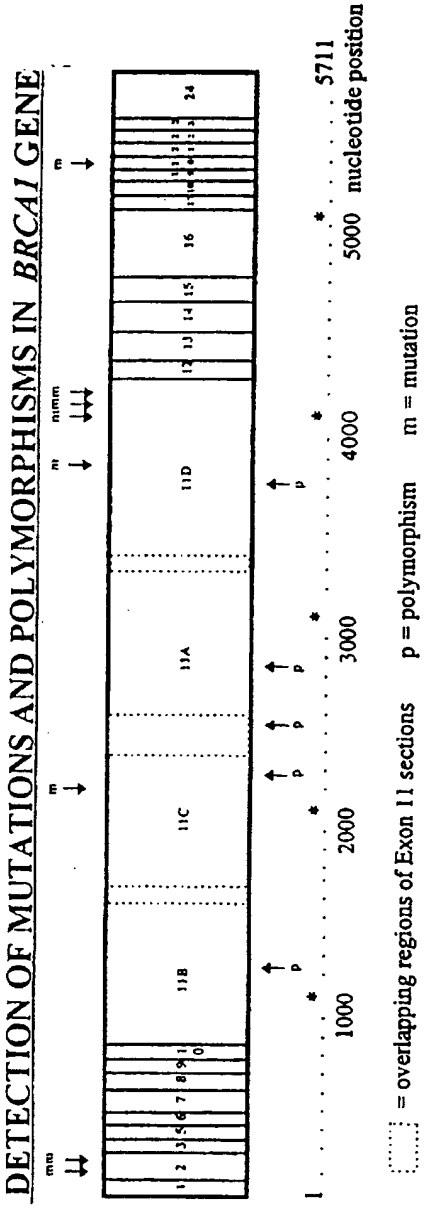

DETECTION OF MUTATIONS AND POLYMORPHISMS IN BRCA1 GENE
1000
2000
3000
4000
5000
5711
nucleotide position
= overlapping regions of Exon 11 sections
p = polymorphism
m = mutation

Figure 10

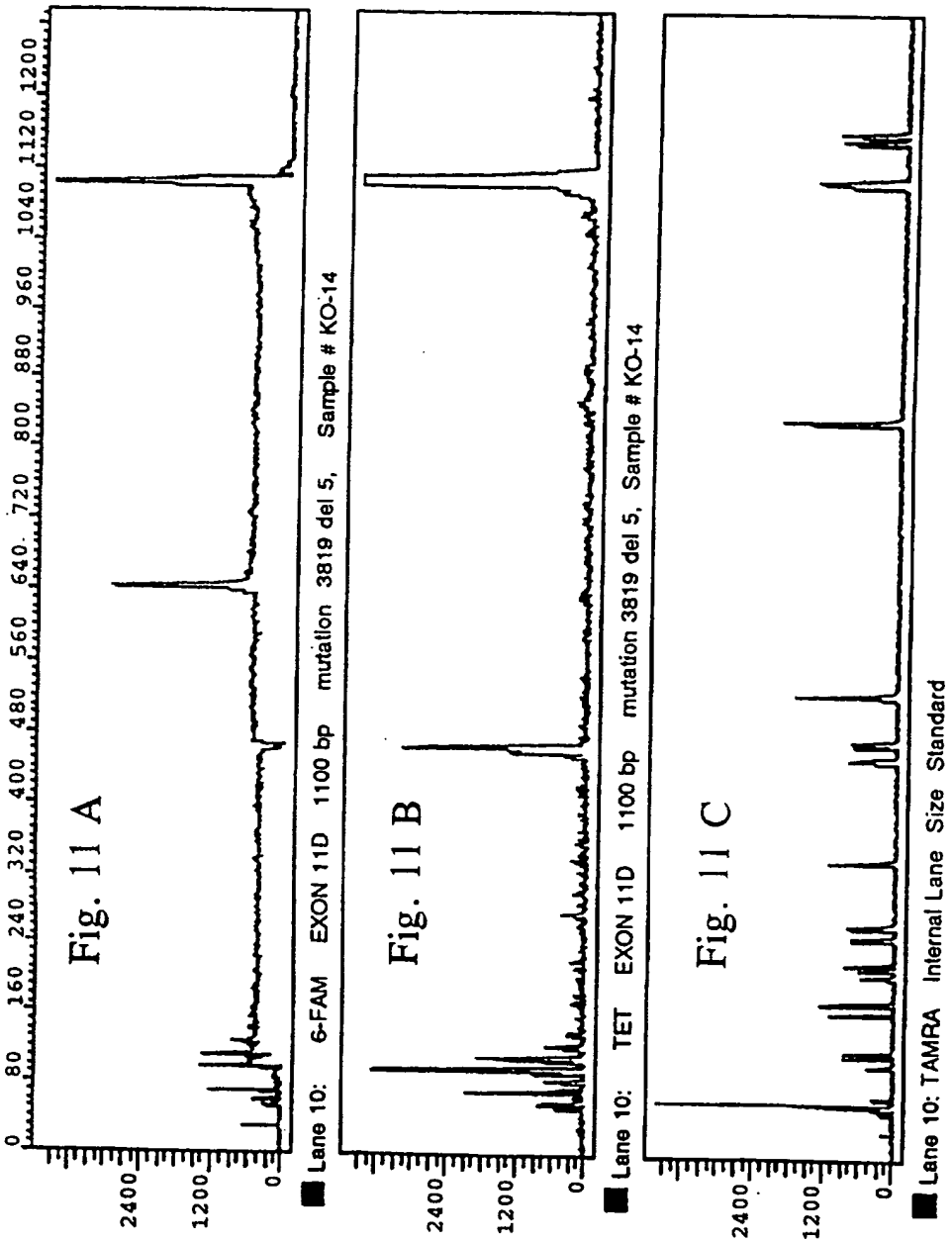


Fig. 11 A
Lane 10: 6-FAM EXON 11D 1100 bp mutation 3819 del 5, Sample # KO-14
Fig. 11 B
Lane 10: TET EXON 11D 1100 bp mutation 3819 del 5, Sample # KO-14
Fig. 11 C
Lane 10: TAMRA Internal Lane Size Standard
0 80 160 240 320 400 480 560 640 720 800 880 960 1040 1120 1200
2400
1200
0

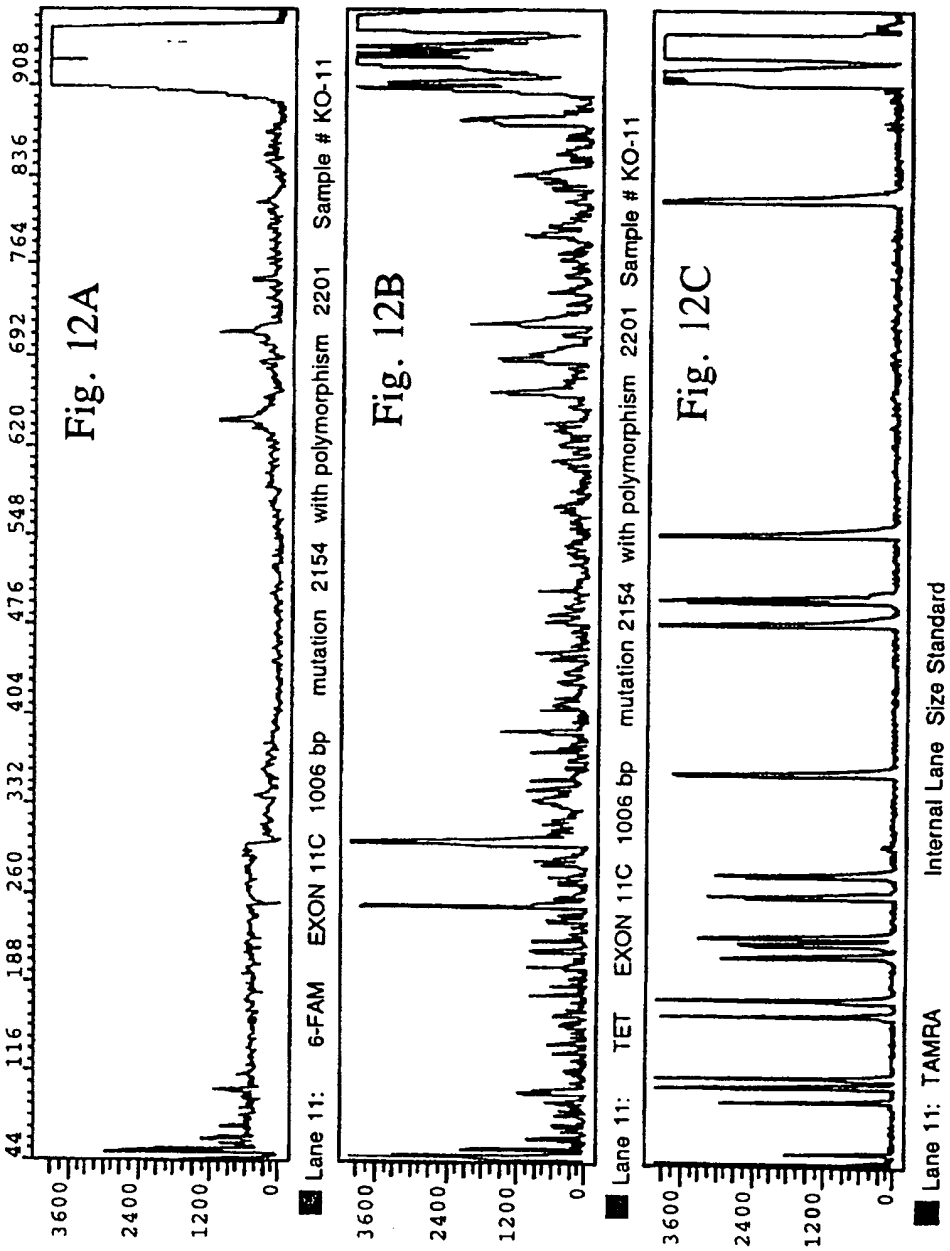


Fig. 12A
Fig. 12B
Fig. 12C
44 116 188 260 332 404 476 548 620 692 764 836 908
3600
2400
1200
0
Lane 11: 6-FAM EXON 11C 1006 bp mutation 2154 with polymorphism 2201 Sample # KO-11
Lane 11: TET EXON 11C 1006 bp mutation 2154 with polymorphism 2201 Sample # KO-11
Lane 11: TAMRA Internal Lane Size Standard

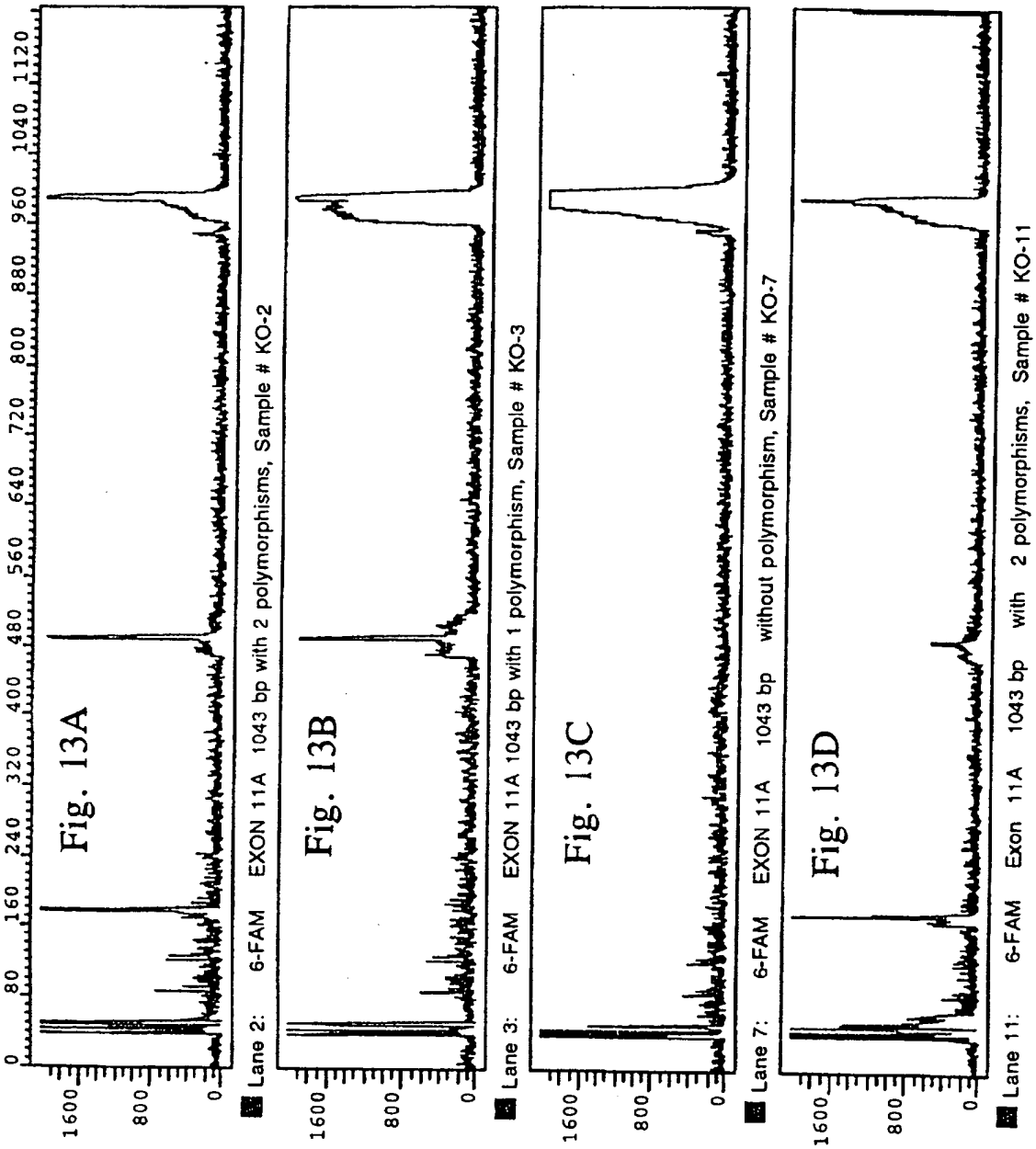


Fig. 13A
Lane 2: 6-FAM EXON 11A 1043 bp with 2 polymorphisms, Sample # KO-2
Fig. 13B
Lane 3: 6-FAM EXON 11A 1043 bp with 1 polymorphism, Sample # KO-3
Fig. 13C
Lane 7: 6-FAM EXON 11A 1043 bp without polymorphism, Sample # KO-7
Fig. 13D
Lane 11: 6-FAM Exon 11A 1043 bp with 2 polymorphisms, Sample # KO-11
0 80 160 240 320 400 480 560 640 720 800 880 960 1040 1120
1600
800
0

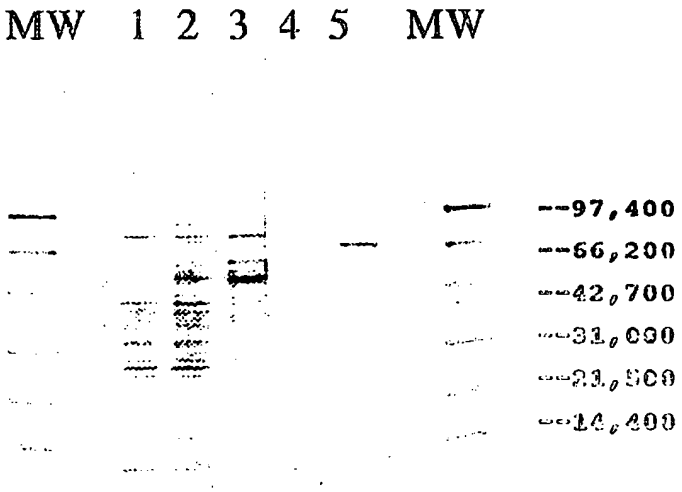


Figure 14

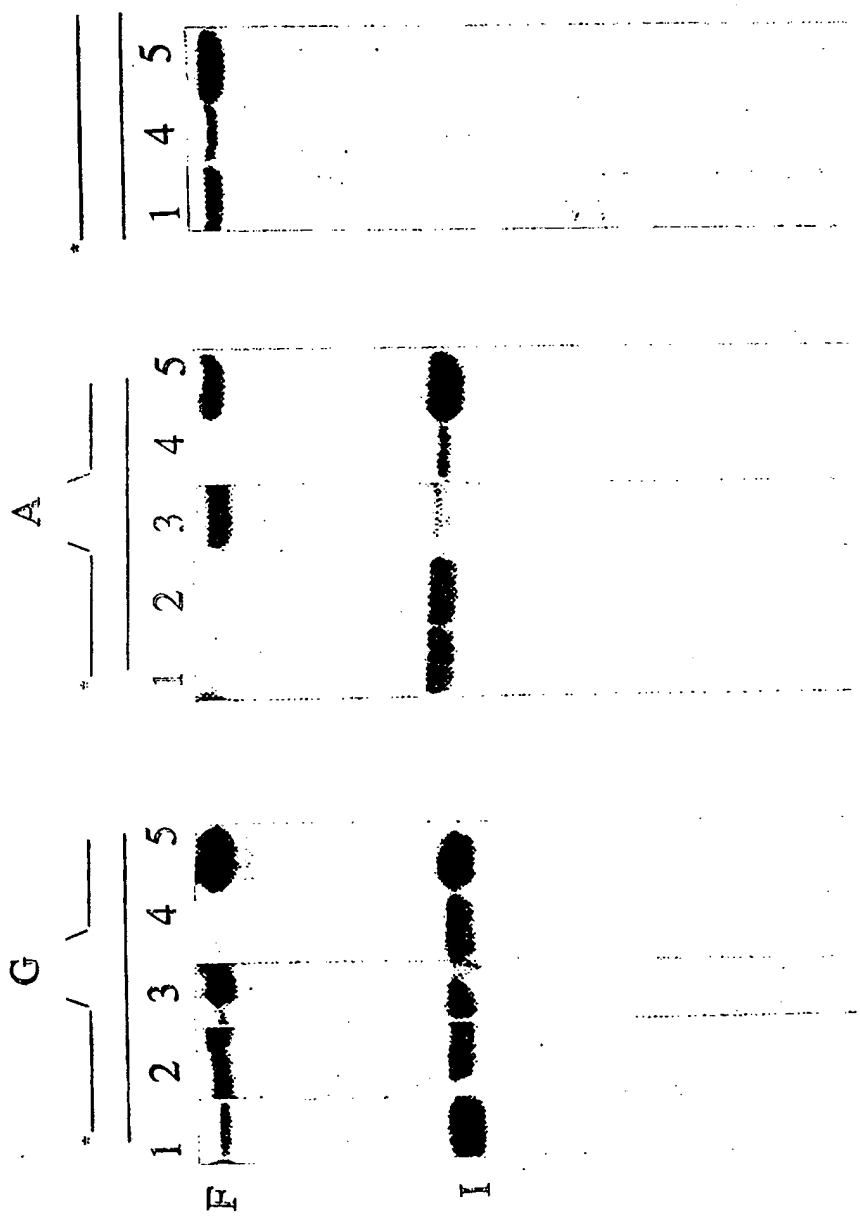


Fig. 15C

Fig. 15B

Fig. 15A

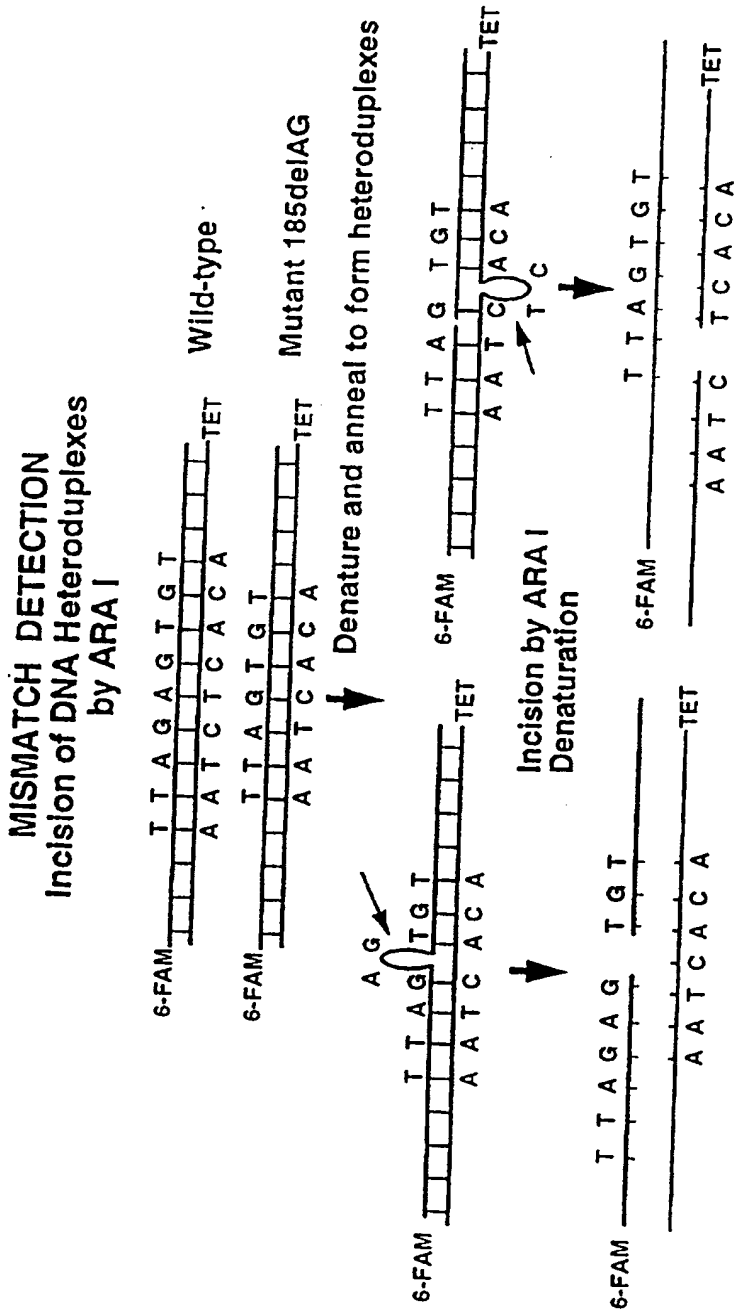

MISMATCH DETECTION
Incision of DNA Heteroduplexes
by ARA I
Wild-type
Mutant 185delAG
Denature and anneal to form heteroduplexes
Incision by ARA I
Denaturation
Gel Electrophoresis
GeneScan
6-FAM
TET
T T A G A G T G T
A A T C T C A C A
T T A G T G T
A A T C A C A
A G
T C

FIGURE 16

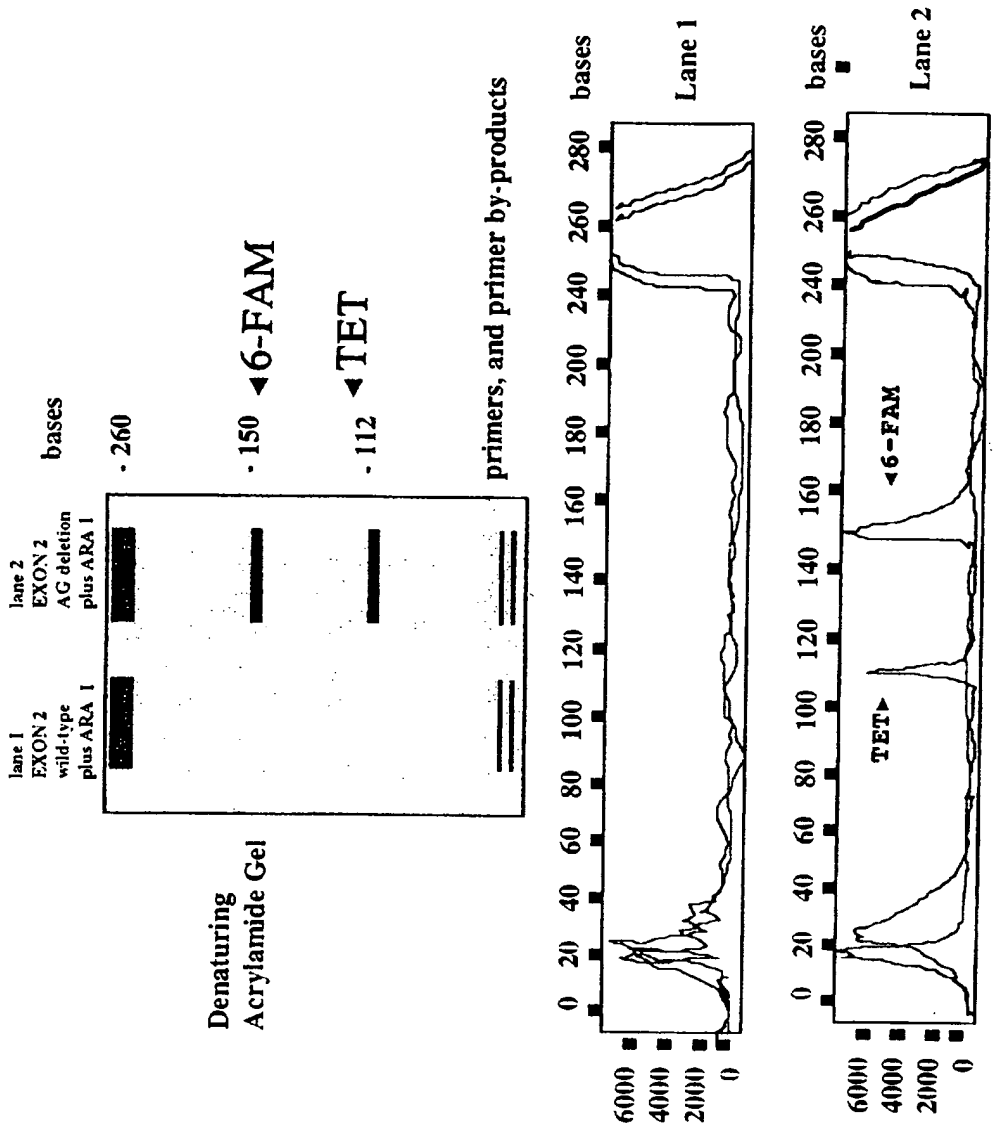

lane 1
EXON 2
wild-type
plus ARA 1
lane 2
EXON 2
AG deletion
plus ARA 1
bases
260
150
6-FAM
112
TET
primers, and primer by-products
Denaturing
Acrylamide Gel
0 20 40 60 80 100 120 140 160 180 200 240 260 280 bases
Lane 1
6000
4000
2000
0
TET
6-FAM
Lane 2

FIGURE 17

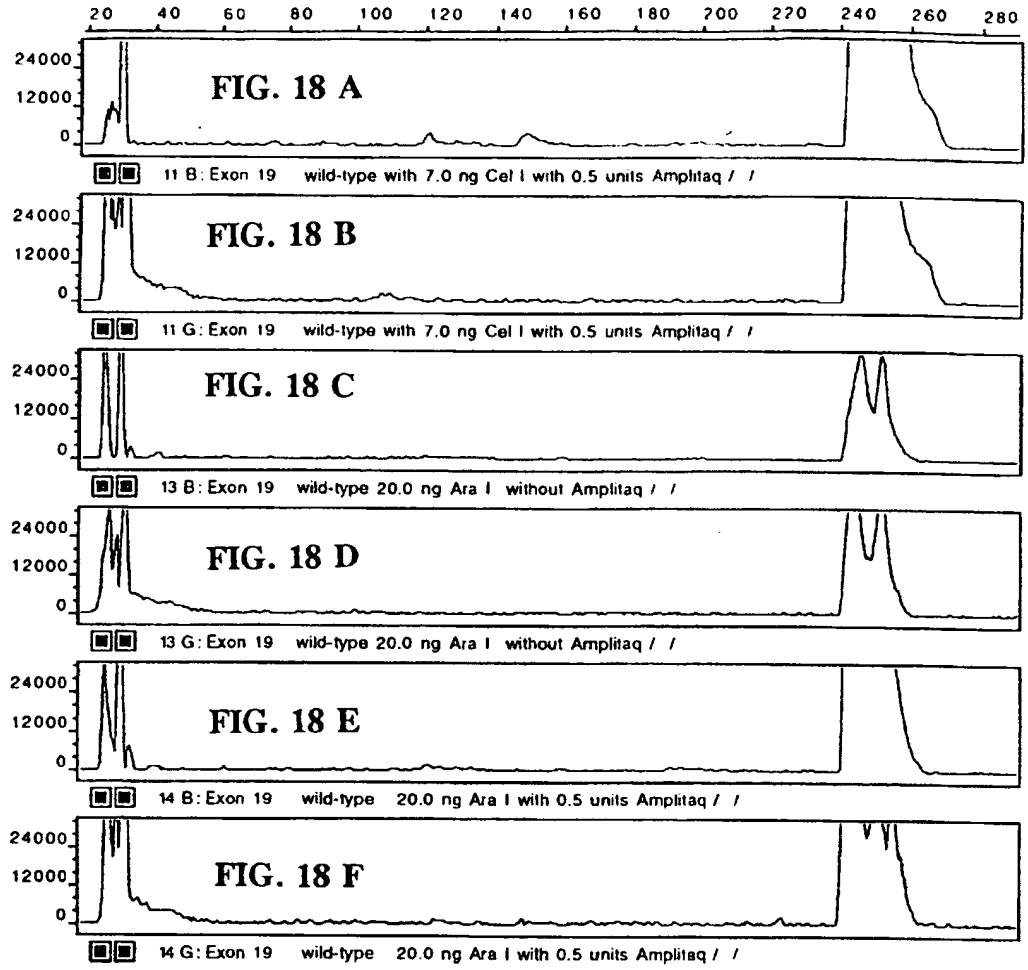


FIG. 18 A

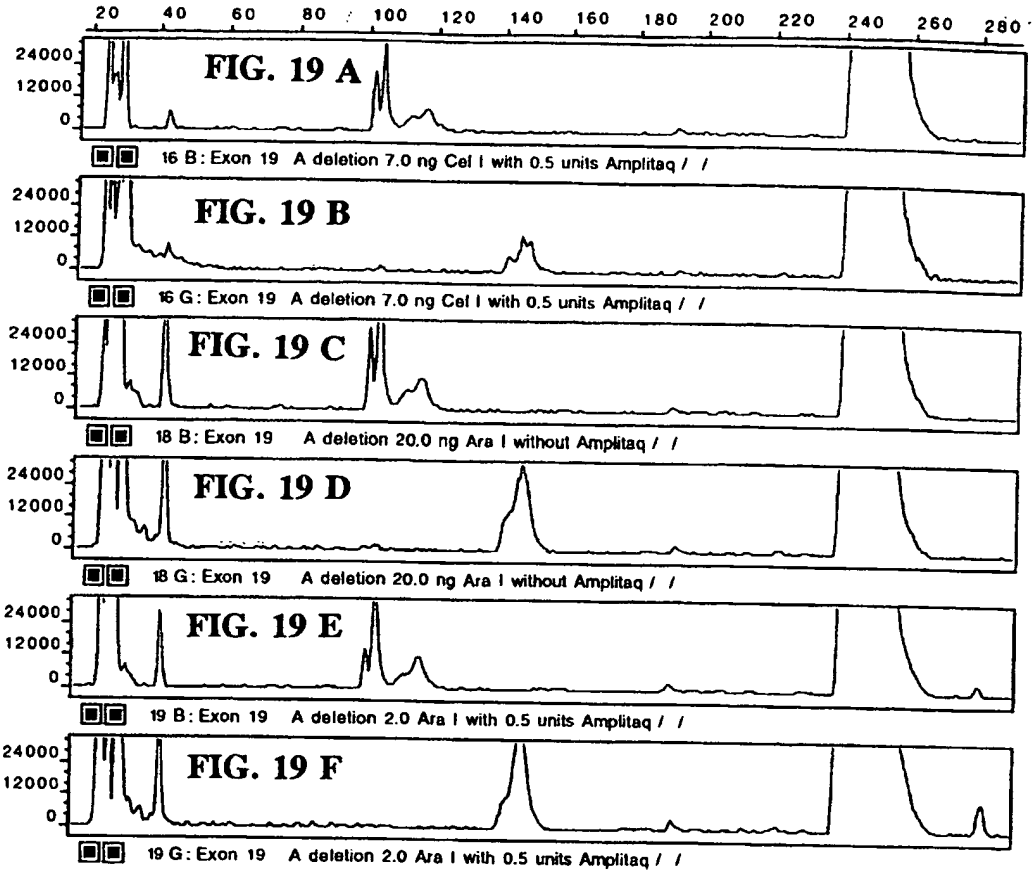
FIG. 18 B

FIG. 18 C

FIG. 18 D

FIG. 18 E

FIG. 18 F


20 40 60 80 100 120 140 160 180 200 220 240 260 280
24000
12000
0
FIG. 19 A
16 B: Exon 19 A deletion 7.0 ng Cel I with 0.5 units Amplitaq / /
FIG. 19 B
16 G: Exon 19 A deletion 7.0 ng Cel I with 0.5 units Amplitaq / /
FIG. 19 C
18 B: Exon 19 A deletion 20.0 ng Ara I without Amplitaq / /
FIG. 19 D
18 G: Exon 19 A deletion 20.0 ng Ara I without Amplitaq / /
FIG. 19 E
19 B: Exon 19 A deletion 2.0 Ara I with 0.5 units Amplitaq / /
FIG. 19 F
19 G: Exon 19 A deletion 2.0 Ara I with 0.5 units Amplitaq / /

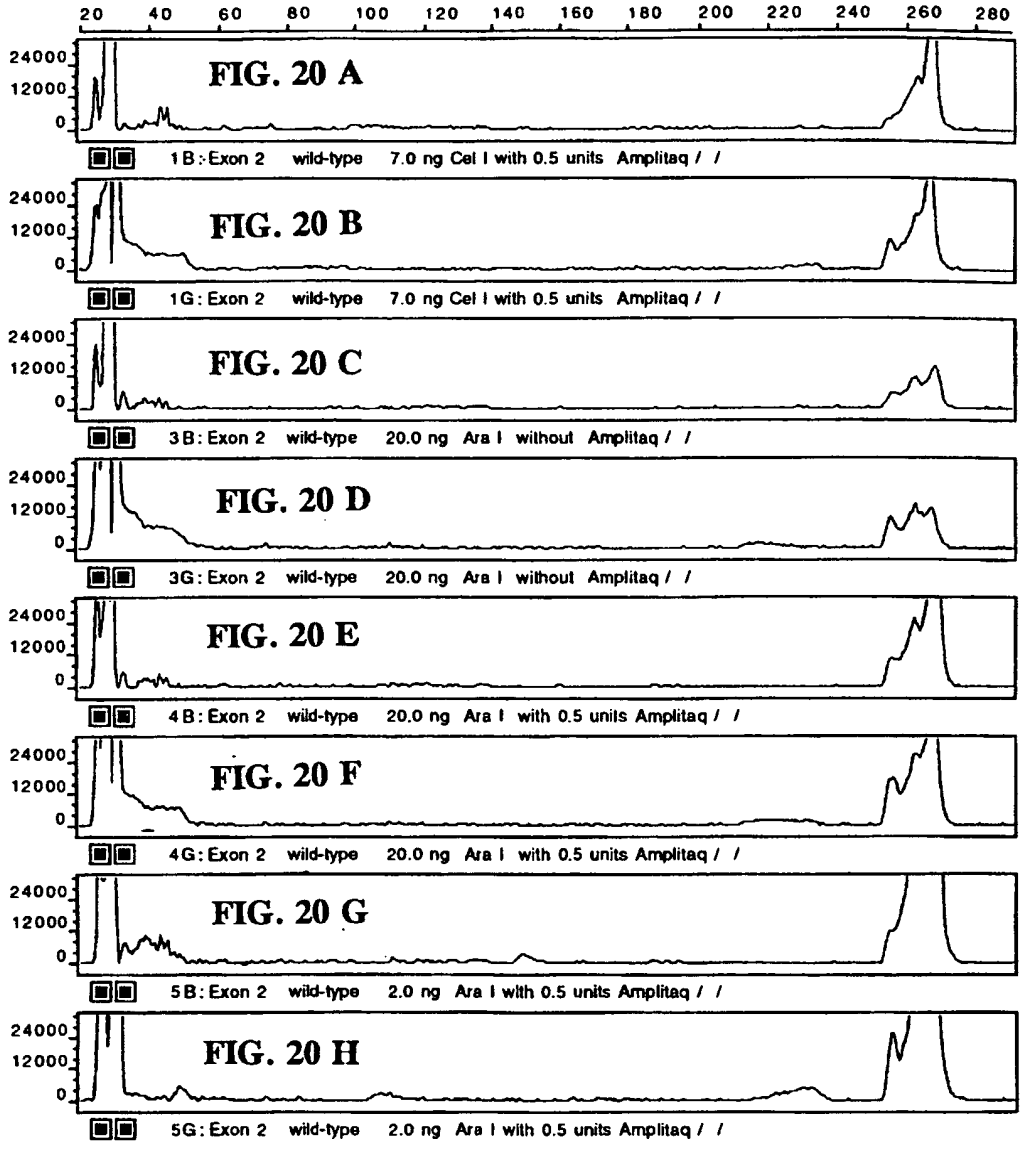


FIG. 20 A

FIG. 20 B

FIG. 20 C

FIG. 20 D

FIG. 20 E

FIG. 20 F

FIG. 20 G

FIG. 20 H

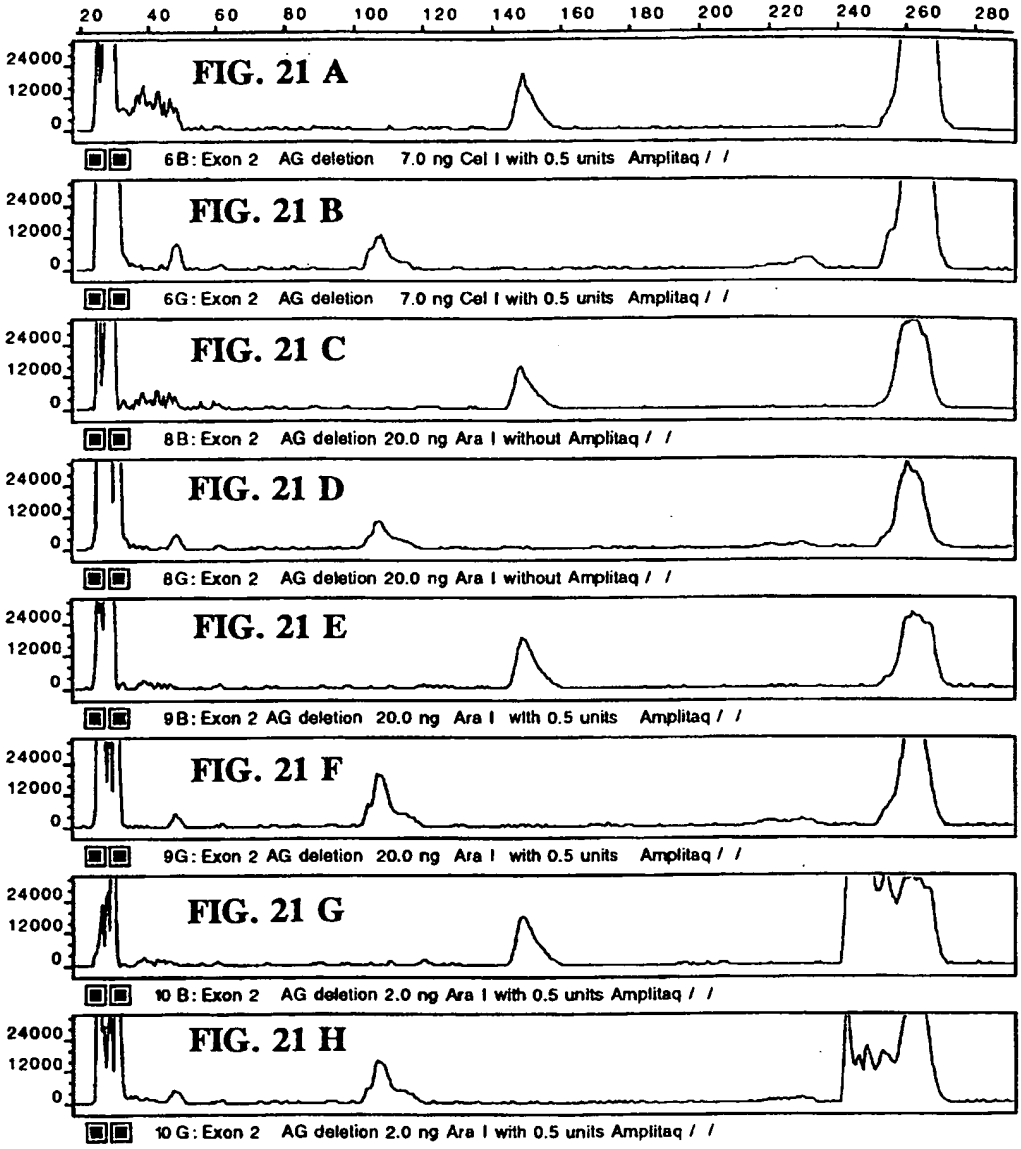

20 40 60 80 100 120 140 160 180 200 220 240 260 280
24000
12000
0
FIG. 21 A
6B: Exon 2 AG deletion 7.0 ng Cel I with 0.5 units Amplitaq / /
FIG. 21 B
6G: Exon 2 AG deletion 7.0 ng Cel I with 0.5 units Amplitaq / /
FIG. 21 C
8B: Exon 2 AG deletion 20.0 ng Ara I without Amplitaq / /
FIG. 21 D
8G: Exon 2 AG deletion 20.0 ng Ara I without Amplitaq / /
FIG. 21 E
9B: Exon 2 AG deletion 20.0 ng Ara I with 0.5 units Amplitaq / /
FIG. 21 F
9G: Exon 2 AG deletion 20.0 ng Ara I with 0.5 units Amplitaq / /
FIG. 21 G
10 B: Exon 2 AG deletion 2.0 ng Ara I with 0.5 units Amplitaq / /
FIG. 21 H
10 G: Exon 2 AG deletion 2.0 ng Ara I with 0.5 units Amplitaq / /

Fig. 22 A

Fig. 22 B

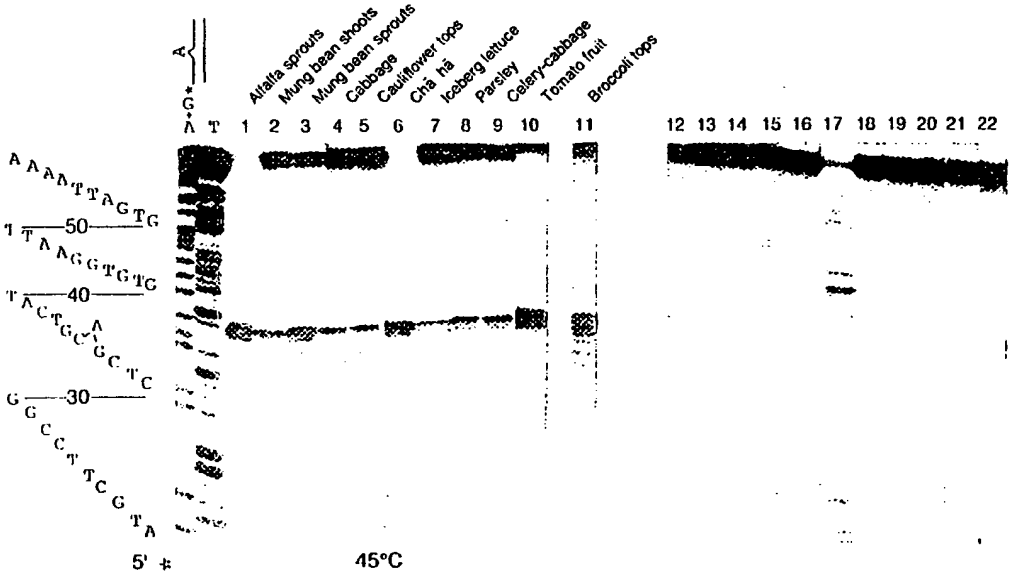
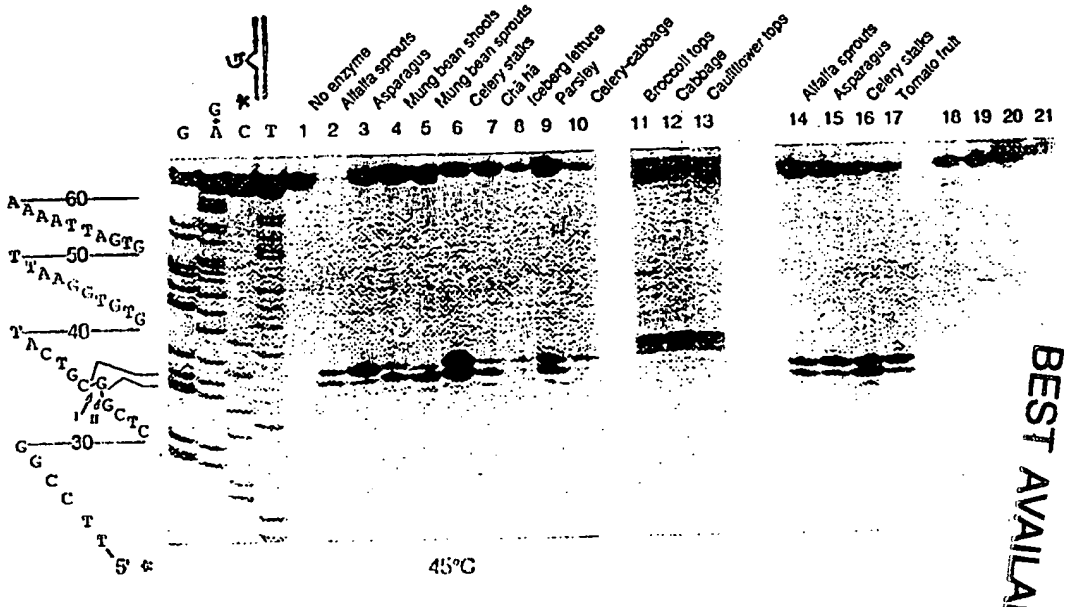


Fig. 23A Fig. 23B



INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/08705

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C12Q 1/68; C12N 9/14, 9/16
US CL :435/6, 195, 196
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 195, 196

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X, P	US 5,571,676 A (SHUBER) 05 November 1996, columns 4-7.	1, 3, 5, 8-14
X --- A	YOUIL et al. Detection of 81 of 81 Known Mouse β-globin Promoter Mutations with T4 Endonuclease VII--The EMC Method. Genomics. 1996, Vol. 32, pages 431-435, see entire document.	1, 3, 5, 8-14 --- 2, 4, 6, 7, 16, 17
X --- A	WO 93/20233 A1 (UNIVERSITY OF MARYLAND AT BALTIMORE) 14 October 1993, see page 20.	1, 3, 5, 8-14 --- 2, 4, 6, 7, 16, 17
X	US 5,459,039 A (MODRICH et al.) 17 October 1995, columns 10 and 13.	1, 3, 5, 8-14

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
28 JULY 1997	28 AUG 1997
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer Amy Atzel, Ph.D. Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/08705

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- A	WO 95/29251 A1 (APPLIED TECHNOLOGY GENETICS CORPORATION) 02 November 1995, see pages 21-33.	1, 3, 5, 8-14 --- 2, 4, 6, 7, 16, 17
X --- A	DODGSON et al. Action of Single-Strand Specific Nucleases on Model DNA Heteroduplexes of Defined Size and Sequence. Biochemistry. 1977, Vol. 16, No. 11, pages 2374-2379, see entire document.	15 --- 16, 17

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/08705

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS, CAPLUS, EMBASE, BIOSIS, MEDLINE, INPADOC, SCISEARCH, LIFESCI
search terms:
resolvase, mung bean, s1, mutY, celery, arabidopsis, Yeung, Modrich, Cotton, mismatch, protection, apium graveolens, nuclease, insertion, deletion, psoralen,

PCT WORL Exp Mail EV335610972US USAN 10/816,459 NIZATION

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68, C07H 21/02, 21/04, C12N 15/00 | A1

(11) International Publication Number: WO 99/22029

(43) International Publication Date: 6 May 1999 (06.05.99)

(21) International Application Number: PCT/US98/23142

(22) International Filing Date: 28 October 1998 (28.10.98)

(30) Priority Data:
60/063,685 28 October 1997 (28.10.97) US

(71) Applicant: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA [US/US]; Los Alamos National Laboratory, LC/BPL, MS D412, Los Alamos, NM 87545 (US).

(71)(72) Applicants and Inventors: NOLAN, John, P. [US/US]; 11 Bundy Road, Santa Fe, NM 87501 (US). WHITE, P., Scott [US/US]; 2942 A Orange Street, Los Alamos, NM 87544 (US). CAI, Hong [CN/US]; 1997 Cumbres Patio, Los Alamos, NM 87544 (US).

(74) Agents: FREUND, Samuel, M. et al.; Los Alamos National Laboratory, LC/BPL, MS D412, Los Alamos, NM 87545 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: DNA BASE MISMATCH DETECTION USING FLOW CYTOMETRY

(57) Abstract

Flow cytometric DNA base mismatch detection. The use of immobilized mismatch–binding protein–coated microspheres to bind fluorescently labeled, mismatch–containing DNA for detection by flow cytometry is described. The genomic DNA of interest is made fluorescent by one of a number of techniques. The resulting fluorescent DNA is melted and allowed to reanneal. If the genomic sample contains a polymorphic site, the reannealed DNA will contain base mismatches. Microspheres bearing immobilized mismatch–binding protein are added to the DNA samples and the mismatch–containing DNA allowed to bind to the mismatch–binding proteins. The sample is then analyzed by flow cytometry to detect the fluorescent DNA bound to the microspheres. The present flow–cytometric method for mutation scanning has several important features: it is a single step, no wash assay, capable of rapid and sensitive analysis. It will allow amplified regions to be rapidly scanned for the presence of polymorphisms to prioritize samples for complete sequencing or to detect mutations in known regions of the genome.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

DNA BASE MISMATCH DETECTION USING FLOW CYTOMETRY

This patent application claims priority from Provisional Patent Application Serial No. 60/063,685 filed on October 28, 1997.

FIELD OF THE INVENTION

The present invention relates generally to the use of flow cytometry and immobilized mismatch binding proteins for the detection of DNA polymorphisms, including nucleotide polymorphisms, insertions, and deletions. This invention was made with government support under Contract No. W-7405-ENG-36 awarded by the U.S. Department of Energy to The Regents of The University of California. The government has certain rights in the invention.

BACKGROUND OF THE INVENTION

As the sequencing of the first human genome approaches completion, significant attention is being given to the evaluation of genetic variation among individuals. An important element of such an effort involves the identification of sites that vary from individual to individual, especially single nucleotide polymorphisms (SNPs). This information could be obtained from direct DNA sequencing of many individuals, but the low throughput and high cost of presently available sequencing procedures make this impractical on a large scale. As a result, considerable effort has been directed towards methods that "scan" DNA sequences for mutations before subjecting them to direct sequencing. These scanning methods have been the subject of frequent reviews. See, e.g., "Slowly But Surely Towards Better Scanning for Mutations," by Richard G.H. Cotton, Trends In Genetics **13**, 43 (1997). For high-throughput SNP discovery, a scanning method should be capable of rapid and parallel analysis with a minimum of sample preparation and handling. Gel-based methods offer some capacity for parallel analyses, but are very slow. High-performance liquid chromatography (HPLC) is often faster, but samples must be analyzed sequentially. Microplate-based assays require time-consuming sample processing, with many incubation and wash steps. However, these assays can

be performed in parallel, as can sample analysis. However, only one or two measurements can be made per microplate well.

A number of these methods are based on detecting differences in physical properties of polymorphism-containing DNA. The single-strand conformation polymorphism method (SSCP) relies on detecting sequence-related differences in the secondary structure of single-stranded DNA as detected by nondenaturing gel electrophoresis. Denaturing gradient-gel electrophoresis (DGGE) detects polymorphisms by the different mobilities of homoduplex versus mismatch-containing heteroduplex double-stranded DNA. The sensitivity of heteroduplex DNA to cleavage by chemicals (hydroxylamine/osmium tetroxide) or enzymes (T4 endonuclease VII or the *E.coli* Must/H/L complex) can be detected by gel electrophoresis.

Although gel electrophoresis is widely available, its use places limitations on analysis speed and throughput and there is significant interest in gel-free methods for mutation scanning. One example that has gained some popularity is the use of immobilized mismatch binding protein (IMP), typically the bacterial mismatch recognition protein MutS, to detect heteroduplex-containing DNA. This approach uses filter binding or microwell plate formats to screen for the presence of mismatch-containing heteroduplex in PCR amplified samples. In "Mutation Detection Using Immobilized Mismatch Binding Protein MutS," by Robert Wagner et al., Nucleic Acids Res. **23**, 3944 (1995), MutS immobilized by binding to nitrocellulose was found to provide a sensitive and accurate mutation detection assay. However, numerous washing steps are required during various steps in the assay. See, also "Allele Identification Using Immobilized Mismatch Binding Protein: Detection And Identification Of Antibiotic Resistant Bacteria And Determination Of Sheep Susceptibility To Scrapie," by P. Debbie et al., Nucleic Acids Res. **25**, 4825 (1997). Recently, the use of an immobilized mismatch cleaving protein to bind heteroduplex DNA for mutation detection has been reported. Endonuclease VII, which cleaves mismatch containing DNA in the presence of Mg^{++}, is immobilized on microtiter plates and used to bind

heteroduplex DNA in the absence of Mg^{++}. See, e.g., "Enzymatic Mutation Detection: Procedure For Screening And Mapping Of Mutations By Immobilization Endonuclease VII," by S. Golz et al., Nucleic Acids Res. **26**, 1132 (1998). Endonuclease VII has the advantage of recognizing all possible mismatches, including C/C mismatches, as well as bulges or loops caused by insertions or deletions. See, also "Detection of 81 Of 81 Known Mouse β-Globin Promoter Mutations With T4 Endonuclease VII-The EMC Method," by R. Youil et al., Genomics **32**, 431 (1996).

Flow cytometry provides multiparameter detection with excellent sensitivity in a homogenous assay format. Multicolor fluorescence detection can be exploited for the simultaneous detection of dozens, or potentially hundreds, of analytes in a single sample. These features enable washless, small volume, multiplexed analyses for high-throughput applications.

Accordingly, it is an object of the present invention to provide a method for detecting SNPs by using flow cytometry to locate mismatch-containing DNA strands attached to mismatch-binding proteins immobilized on microspheres.

Another object of the invention is to provide a high-throughput, washless, small volume method for detecting SNPs using flow cytometry.

Additional objects, advantages and novel features of the invention will be set forth in part in the description which follows, and in part will become apparent to those skilled in the art upon examination of the following or may be learned by practice of the invention. The objects and advantages of the invention may be realized and attained by means of the instrumentalities and combinations particularly pointed out in the appended claims.

SUMMARY OF THE INVENTION

To achieve the foregoing and other objects, and in accordance with the purposes of the present invention, as embodied and broadly described herein, the method for the detection of DNA polymorphisms hereof may comprise in combination: immobilizing a mismatch-binding protein on microspheres; preparing fluorescently labeled copies of the DNA to be interrogated; and flowing

the microspheres through a flow cytometer, wherein the binding of the fluorescent DNA to the microspheres is measured.

Preferably, the mismatch binding protein is the bacterial mismatch binding protein MutS or a homologue thereof.

Benefits and advantages of the present invention include a high-throughput, small volume, washless method for detecting SNPs in DNA.

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated in and form a part of the specification, illustrate the embodiments of the present invention and, together with the description, serve to explain the principles of the invention. In the drawings:

FIGURE 1a is a schematic representation of molecules of MutS, which bind to SNP locations in DNA, immobilized on a microsphere in the presence of DNA strands having SNPs and strands of normal DNA, while FIG. 1b is a schematic representation of binding to the microsphere of the DNA strands which have SNPs as a result of the mismatch-binding protein immobilized thereon, normal DNA strands remaining in solution, whereby DNA having SNPs can be detected using flow cytometry.

FIGURE 2 is a graph of the fluorescence, as measured by flow cytometry, of immobilized mismatch-binding protein-coated microspheres as a function of the concentration of mismatch-containing DNA (open circles) or non-mismatch-containing DNA (black circles) after 30 min. of co-incubation.

DETAILED DESCRIPTION OF THE INVENTION

Briefly, the present invention includes the use of immobilized mismatch-binding protein-coated microspheres to bind fluorescently labeled, mismatch-containing DNA for detection by flow cytometry. The genomic DNA of interest is amplified by polymerase chain reaction (PCR) using fluorescently labeled nucleotide triphosphates. The resulting fluorescent DNA is melted and allowed to reanneal. If the genomic sample contains a polymorphic site, the reannealed DNA will contain base mismatches. Microspheres bearing immobilized

mismatch-binding protein are added to the DNA samples and the mismatch-containing DNA allowed to bind to the mismatch-binding proteins. The sample is then analyzed by flow cytometry to detect the fluorescent DNA bound to the microspheres. Alternatively, DNA may be incubated with soluble mismatch-binding protein, followed by immobilization of the DNA-mismatch-binding protein complex onto microspheres for analysis using flow cytometry. By adapting the assay to flow cytometry techniques, existing SNP detection methods using immobilized mismatch-binding protein have been improved. The present flow-cytometric method for mutation scanning has several important features. It is a single step, no wash assay, capable of rapid and sensitive analysis and, as such, will allow amplified regions to be rapidly scanned for the presence of polymorphisms to prioritize samples for complete sequencing or to detect mutations in known regions of the genome.

Reference will now be made to the present preferred embodiments of the invention, examples of which are shown in the accompanying drawings. Turning now to the Figures, Fig. 1a is a schematic representation of molecules of MutS, which bind to SNP locations in DNA, immobilized on a microsphere in the presence of DNA strands having SNPs and strands of normal DNA, while FIG. 1b is a schematic representation of the binding to the microsphere of the DNA strands which have SNPs as a result of the mismatch-binding protein immobilized thereon. The normal DNA strands remain in solution. DNA having SNPs can then be detected using flow cytometry.

Mismatch-binding proteins include the bacterial mismatch-binding protein MutS, homologues thereof, the T4 bacteriophage enzyme endonuclease VII, one of its functional homologues, or any other protein that recognizes DNA base pair mismatches and which can be immobilized on microspheres by physical adsorption or by the use of an affinity tag which binds to an affinity partner immobilized on the microspheres. Useful affinity tags and binding partners are listed in the TABLE.

TABLE. Affinity Tags and Binding Partners

Affinity Tag	Binding Partner
Biotin	avidin/streptavidin
Hexahistidine	Metal chelates
Glutathione-S-transferase	Glutathione
Maltose binding protein	Maltose
Cellulose binding domain	Cellulose
Protein A	IgG

The microspheres are composed of polystyrene, cellulose, agarose, or other suitable material, and bear one of the affinity tag binding partners listed in the TABLE.

Fluorescently labeled DNA is prepared by amplification using the polymerase chain reaction with fluorescently labeled primers or fluorescently labeled nucleotide triphosphates, or by post-amplification staining with a fluorescent DNA binding dye. Alternatively, the DNA can be labeled using biotinylated nucleotide triphosphates in a polymerase chain reaction amplification process, followed by binding with fluorescently labeled avidin, after the DNA has been fixed onto microspheres by the mismatch-binding protein coating thereon.

Having generally described the present invention, the following EXAMPLE is intended to provide more specific details thereof.

EXAMPLE

Microspheres bearing immobilized mismatch-binding protein are prepared by adsorbing one nanomole of mismatch-binding protein on one hundred million microspheres, each ten micrometers in diameter. The incubation occurs efficiently in a total volume of one milliliter of Tris buffer (50 mM Tris, 100 mM NaCl, pH 8.0) chilled for 12 hours using ice, followed by washing which includes two cycles of centrifugation and resuspension. The DNA samples to be tested are amplified by polymerase chain reaction (PCR) using a high fidelity thermostable DNA polymerase, such as *Pfu*, to optimize amplification fidelity.

Fluorescein-labeled dUTP is included in the amplification reaction such that the amplified DNA contains the fluorescent label. The amplified DNA is then melted and allowed to reanneal. Microspheres bearing immobilized mismatch-binding protein are added (one hundred thousand to a total volume of twenty microliters), and mismatch-containing double-stranded DNA is allowed to bind to the microspheres. The samples are then diluted with buffer to a final volume of five hundred microliters and the microsphere-associated fluorescence is measured in a conventional flow cytometer. Figure 2 is a graph of the fluorescence, as measured by flow cytometry, of immobilized mismatch-binding protein-coated microspheres as a function of the concentration of mismatch-containing DNA (open circles) or non-mismatch-containing DNA (black circles) after 30 min. of co-incubation. As can be seen from Fig. 2, at DNA concentrations greater than 100 nM, the specific binding of the mismatch-containing DNA is clearly distinguishable from the background of normal DNA.

The foregoing description of the invention has been presented for purposes of illustration and description and is not intended to be exhaustive or to limit the invention to the precise form disclosed, and obviously many modifications and variations are possible in light of the above teaching. The embodiments were chosen and described in order to best explain the principles of the invention and its practical application to thereby enable others skilled in the art to best utilize the invention in various embodiments and with various modifications as are suited to the particular use contemplated. It is intended that the scope of the invention be defined by the claims appended hereto.

WHAT IS CLAIMED IS:

1. A method for detecting mismatch-containing DNA, which comprises the steps of:

(a) fluorescently labeling the DNA under investigation;

(b) preparing microspheres bearing immobilized mismatch-binding protein;

(c) incubating the fluorescently labeled sample of DNA with the microspheres having mismatch-binding protein immobilized thereon, thereby binding mismatch-containing, fluorescently labeled DNA to the microspheres; and

(d) detecting the mismatch-containing, fluorescently labeled DNA bound to the microspheres by flow cytometry.

2. The method for detecting mismatch-containing DNA as described in claim 1, wherein said step of fluorescently labeling the DNA under investigation is achieved by polymerase chain reaction amplification using a fluorescently labeled primer.

3. The method for detecting mismatch-containing DNA as described in claim 1, wherein said step of fluorescently labeling the DNA under investigation is achieved by polymerase chain reaction amplification in the presence of fluorescent deoxynucleotide triphosphates.

4. The method for detecting mismatch-containing DNA as described in claim 1, wherein said step of fluorescently labeling the DNA under investigation is achieved by using a DNA-binding fluorophore.

5. The method for detecting mismatch-containing DNA as described in claim 1, wherein the mismatch-binding protein is selected from the group consisting of *Escherichia coli* MutS and T4 endonuclease VII.

6. The method for detecting mismatch-containing DNA as described in claim 1, wherein the microspheres are composed from materials selected from the group consisting of polystyrene and cellulose.

7. The method for detecting mismatch-containing DNA as described in claim 1, wherein the mismatch-binding protein is immobilized on the microspheres by physical adsorption.

8. The method for detecting mismatch-containing DNA as described in claim 1, wherein the mismatch-binding protein is immobilized on the microspheres by covalent attachment.

9. The method for detecting mismatch-containing DNA as described in claim 1, wherein the mismatch-binding protein is immobilized on the microsphere by an affinity tag pair selected from the group consisting of avidin-biotin, hexahistidine-metal chelate, and glutathione-S-transferase-glutathione.

10. A method for detecting mismatch-containing DNA which comprises the steps of:

(a) amplifying the DNA under investigation using the polymerase chain reaction in the presence of biotinylated deoxynucleotide triphosphates, thereby producing biotinylated DNA;

(b) preparing microspheres bearing immobilized mismatch-binding protein;

(c) incubating the biotinylated DNA with microspheres bearing the immobilized mismatch-binding protein, thereby binding mismatch-containing, biotinylated DNA to the microspheres;

(d) incubating the microspheres having mismatch-containing, biotinylated DNA bound thereto with fluorescently labeled avidin, thereby producing mismatch-containing, fluorescently labeled DNA bound to the microspheres; and

(e) detecting the mismatch-containing, fluorescently labeled DNA bound to the microspheres by flow cytometry.

11. The method for detecting mismatch-containing DNA as described in claim 10, wherein the mismatch-binding protein is selected from the group consisting of *Escherichia coli* MutS and T4 endonuclease VII.

12. The method for detecting mismatch-containing DNA as described in claim 10, wherein the microspheres are composed from materials selected from the group consisting of polystyrene and cellulose.

13. The method for detecting mismatch-containing DNA as described in claim 10, wherein the mismatch-binding protein is immobilized on the microspheres by physical adsorption.

14. The method for detecting mismatch-containing DNA as described in claim 10, wherein the mismatch-binding protein is immobilized on the microspheres by covalent attachment.

15. The method for detecting mismatch-containing DNA as described in claim 10, wherein the mismatch-binding protein is immobilized on the microsphere by an affinity tag pair selected from the group consisting of avidin-biotin, hexahistidine-metal chelate, and glutathione-S-transferase-glutathione.

16. A method for detecting mismatch-containing DNA, which comprises the steps of:

(a) fluorescently labeling the DNA under investigation;

(b) incubating soluble mismatch-binding protein with the fluorescently labeled DNA, forming thereby a mismatch-binding protein-fluorescently labeled DNA complex;

(b) preparing microspheres adapted to immobilize mismatch-binding protein;

(c) incubating the complex with the microspheres adapted to immobilize mismatch-binding protein, thereby binding mismatch-containing, fluorescently labeled DNA to the microspheres; and

(d) detecting the mismatch-containing, fluorescently labeled DNA bound to the microspheres by flow cytometry.

17. The method for detecting mismatch-containing DNA as described in claim 16, wherein the mismatch-binding protein is selected from the group consisting of *Escherichia coli* MutS and T4 endonuclease VII.

18. The method for detecting mismatch-containing DNA as described in claim 16, wherein the microspheres are composed from materials selected from the group consisting of polystyrene and cellulose.

19. The method for detecting mismatch-containing DNA as described in claim 16, wherein the mismatch-binding protein is immobilized on the microspheres by physical adsorption.

20. The method for detecting mismatch-containing DNA as described in claim 16, wherein the mismatch-binding protein is immobilized on the microspheres by covalent attachment.

21. The method for detecting mismatch-containing DNA as described in claim 16, wherein the mismatch-binding protein is immobilized on the microsphere by an affinity tag pair selected from the group consisting of avidin-biotin, hexahistidine-metal chelate, and glutathione-S-transferase-glutathione.

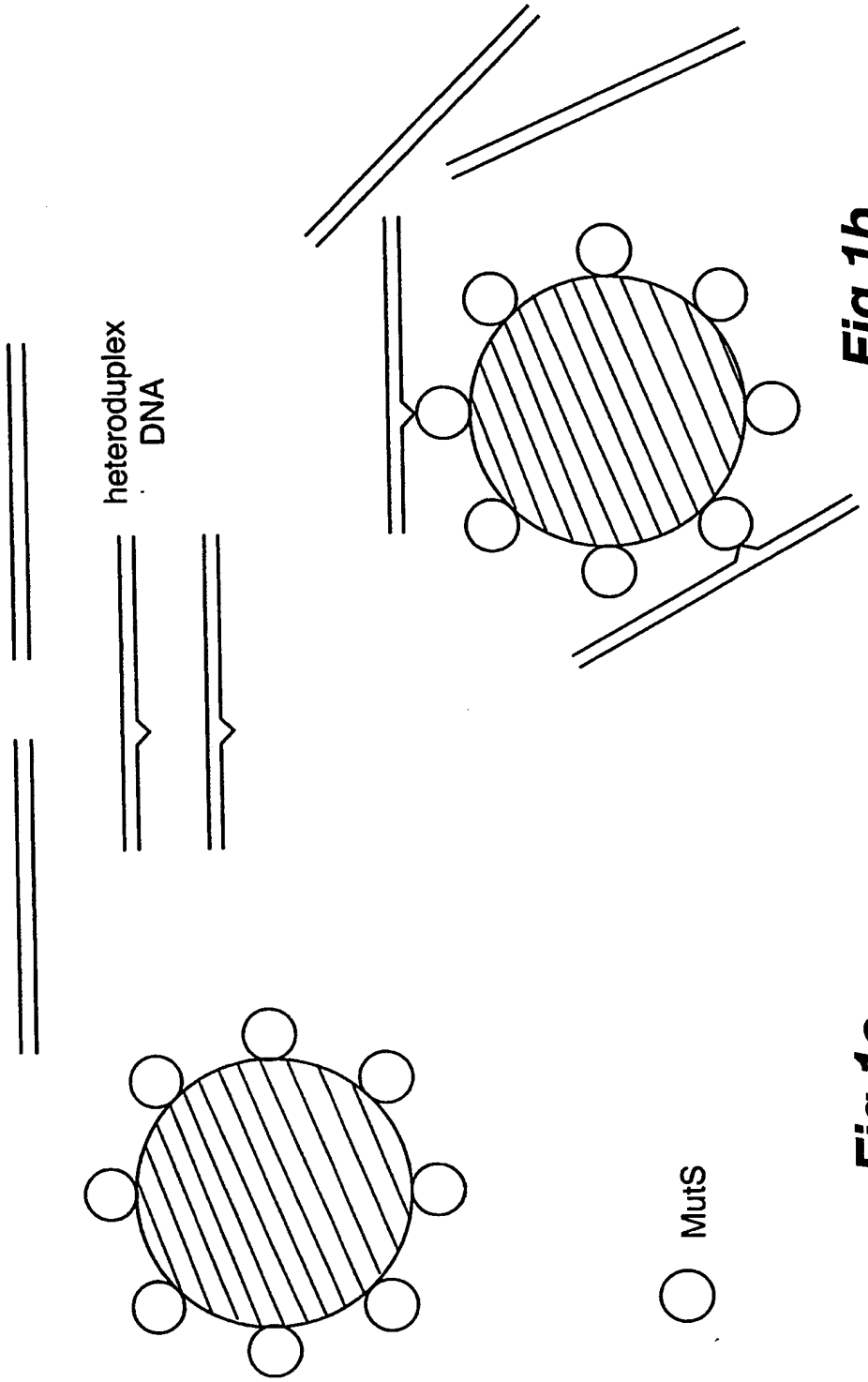


Fig.1a

Fig.1b

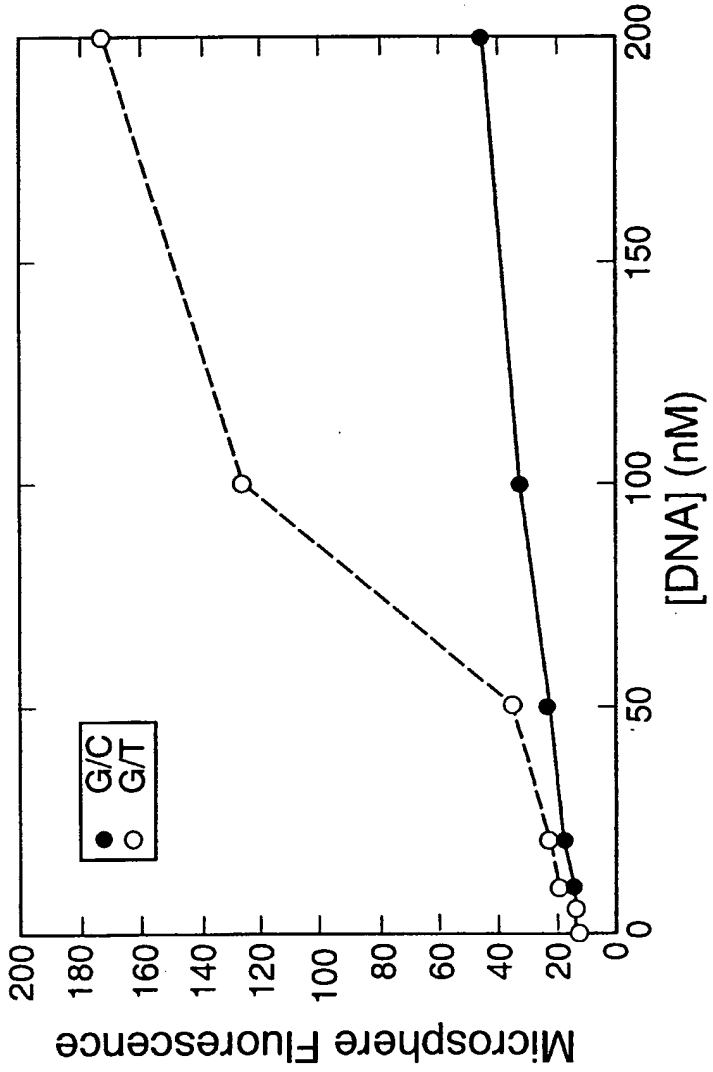

Microsphere Fluorescence
200
180
160
140
120
100
80
60
40
20
0
0
50
100
150
200
[DNA] (nM)
G/C
G/T

Fig. 2

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/23142

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C12Q 1/68; C07H 21/02, 21/04; C12N 15/00

US CL :435/6; 536/23.1, 24.3; 935/76, 77, 78

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6; 536/23.1, 24.3; 935/76, 77, 78

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y,P	US 5,736,330 A(FULTON et al) 07 April 1998, see the entire document.	1-21
Y	WAGNER et al. Mutation Detection using Immobilized Mismatch Binding Protein (MutS). Nucleic Acids Research. 1995. Vol. 23. No. 19. pages 3944-3948, see the entire document.	1-21
Y	WRIGHT et al. CASTing for Multicomponent DNA-Binding Complexes. Trends in Biochemical Sciences (TIBS). March 1993. Vol. 18. pages 77-80, see the entire document.	1-21
Y	LEW et al. Affinity Selection of Polymerase Chain Reaction Products by DNA-Binding Proteins. Methods in Enzymology. 1993. Vol. 218. pages 526-534, see the entire document.	1-21

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 14 DECEMBER 1998	Date of mailing of the international search report 08 FEB 1999
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer ETHAN WHISENANT Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/23142

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	THIESEN et al. Target Detection Assay (TDA): a versatile procedure to determine DNA binding sites as demonstrated on SP1 protein. Nucleic Acids Research. 1990. Vol. 18. No. 11. pages 3203-3209.	1-21
Y	US 5,437,976 A (UTERMOHLEN et al.) 01 August 1995, see the entire document.	1-21
A	US 5,683,877 (LU-CHANG et al) 04 November 1997, see the entire document.	1-21
A	US 5,683,877 A (LU-CHANG et al.) 04 November 1997, see the entire document.	1-21

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/23142

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

USPATFULL, EUROPATFULL, BIOSIS, CAPLUS, MEDLINE

search terms: flow cytometry and (mismatch or mutation) and DNA binding protein

THIS PAGE BLANK (USPTO)